University of Zurich[UZH]

Department of Informatics

Sensorimotor Contingencies in Artificial Intelligence and Robotics

Dissertation submitted to the Faculty of Economics,
Business Administration and Information Technology
of the University of Zurich

to obtain the degree of
Doktor der Wissenschaften, Dr. sc.
(corresponds to Doctor of Science, PhD)

presented by

Nico M. Schmidt

from Germany

approved in April 2015

at the request of
Prof. Dr. Rolf Pfeifer
Prof. Dr. Daniel Polani

University of Zurich[UZH]

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, April 1, 2015

Chairwoman of the Doctoral Board: Prof. Dr. Elaine M. Huang

Acknowledgements

First of all, I would like to thank my supervisor, Prof. Rolf Pfeifer, for giving me the opportunity to join the Artificial Intelligence Laboratory and to support my research that led to this thesis. Further, I would like to thank Prof. Peter König, who provided much useful and constructive input in numerous discussions and who allowed me to spend some months at the "Neurobiopsychology Lab" of the University of Osnabrück, Germany. I am also grateful to Prof. Danica Kragic for inviting me to visit the "Computer Vision and Active Perception Lab" at the Swedish Royal Institute of Technology (KTH). Finally, my appreciation goes to Prof. Daniel Polani for co-reviewing this thesis and to Prof. Renato Pajarola for joining the Ph.D. committee.

There are several people who contributed directly or indirectly to the realization of this work. Special thanks go to all my colleagues and friends at the AI-Lab for the excellent cooperative and cheerful research spirit. I found myself extremely lucky to be able to work with Matej Hoffmann and Kohei Nakajima, my mentors and friends who were always ready for inspiring discussions and sharing of research experience. Thanks to Hugo Gravato Marques for easing EU-project management. Thanks also to Naveen Kuppuswamy for the great time inside and outside the lab. Many thanks to Tamar Tolcachier-Sommer and Nathan Labhart for their precious company and for making administrative procedures smooth. Further thanks go to Matthias Baumgartner, Niklas Wilming, Zhao Qian, Gert Kootstra and Alexander Maye. Last but not least I wish to thank my wife Lena for standing with me and encouraging me in all my ventures.

Abstract

The *pragmatic turn* is a recent trend in cognitive science in which more and more emphasis is put on the action as playing a constitutive role for cognitive processing. The sensorimotor contingencies theory (SMCT) goes along with this trend in putting the interaction process of an agent with its environment at the heart of perception and cognition. According to this theory, perceptions are emerging from the knowledge about the relations between sensory and motor signals that are experienced while enacting the world. Those relations, the sensorimotor contingencies (SMC), define how sensory input changes as an effect of own movements and are believed to be a key building block of lower-level perceptional and possibly higher-level cognitive processes.

In this thesis, we investigate the implications of adopting this perspective for artificial intelligence and robotics. We discuss learning and control in artificial agents based on SMCs and how it is different from classical approaches. This discussion is approached by investigating several aspects of SMCs for artificial intelligence. First, we use information-theoretic tools to identify SMCs in an operating robot and discuss what the robot can learn about itself and its environment by learning the SMCs that are inherent to its brain-body-environment system. Second, we show how captured SMCs can be utilized for learning of tasks-oriented behavior by employing reinforcement learning methodology. Third, we draw parallels with human models of SMCs and discuss how they could translate to control of artificial systems.

Our results show that by observing its SMCs, an agent can learn a lot from its sensorimotor stream during interaction with the environment. Importantly, a minimal amount of externally provided knowledge is required when letting the agent self-explore its sensorimotor space through (initially) random movements. The learned SMCs can provide the agent with certain knowledge about its morphology, the type of its sensors and their topological arrangement. It can furthermore identify and discriminate environmental structures such as objects. By providing a notion of intentions it can finally learn which actions affect its internal values or rewards in which way, in order to select reward-maximizing actions.

Zusammenfassung

Ein entscheidender Wechsel ist jüngst in den Kognitionswissenschaften zu beobachten, der sogenannte *pragmatic turn*, in dem der Handlung mehr und mehr Bedeutung als konstituierendes Element für kognitive Prozesse beigemessen wird. Die *sensorimotor contingency theory* (SMCT) geht mit diesem Trend, indem sie den Interaktionsprozess eines Agenten mit seiner Umwelt in den Mittelpunkt von Wahrnehmung und Kognition stellt. Nach dieser Theorie entstehen Wahrnehmungen aus dem Wissen über die Relationen zwischen sensorischen und motorischen Signalen, die während der Interaktion mit der Welt erfahren werden. Diese Relationen, die *sensorimotor contingencies* (SMC) genannt werden, definieren, wie sich sensorischer Input als Ergebnis von eigenen Bewegungen ändert und werden als Bausteine von Wahrnehmungen niedriger Ebenen, sowie von potentiell höheren kognitiven Prozessen betrachtet.

In dieser Arbeit untersuchen wir die Implikationen dieser Ansichtsweise für die Künstliche Intelligenz und Robotik. Wir behandeln auf SMCs basierendes Lernen und Steuerung in künstlichen Agenten und wie diese sich von klassischen Ansätzen unterscheiden. Wir nähern uns der Diskussion durch Untersuchung von verschiedenen Aspekten von SMCs für die Künstliche Intelligenz. Erstens bedienen wir uns informationstheoretischer Werkzeuge um SMCs in einem operierenden Roboter zu identifizieren und erörtern, was der Roboter über sich selbst und seine Umwelt lernen kann, indem er die SMCs seines *brain-body-environment* Systems erlernt. Zweitens zeigen wir auf, wie durch Verwendung von Methoden des Verstärkenden Lernens, erfasste SMCs zum Lernen von anwendungsorientierten Verhalten benutzt werden können. Drittens ziehen wir Parallelen zu Modellen von menschlichen SMCs und erörtern, wie diese auf die Steuerung von künstlichen Systemen übersetzt werden können.

Unsere Resultate zeigen, dass ein Agent durch Beobachten der eigenen SMCs während der Interaktion mit der Umgebung eine Menge aus seinem sensomotorischen Signalströmen lernen kann. Interessanterweise ist ein nur minimales extern bereitgestelltes Wissen notwendig, wenn der Agent seinen sensomotorischen Zustandsraum durch anfangs zufällige Bewegungen selbst erkundet. Die gelernten SMCs können dem Agenten bestimmtes Wissen über seine Morphologie, die Art seiner Sensoren und ihrer topologischen Anordnung vermitteln. Des weiteren kann er Strukturen seiner Umwelt als Objekte identifizieren und zwischen ihnen unterscheiden. Indem er mit Intention versehen wird, kann er schliesslich lernen, welche Handlungen sein internes Werte- oder Belohnungssystem auf welche Weise beeinflussen, um somit belohnungsmaximierende Handlungen ausführen zu können.

Contents

Chapter 1

Introduction

1.1 Motivation: embodied sensorimotor learning as opposed to symbol processing

How can artificial agents acquire cognitive abilities that enable them to survive in real-world environments, where humans and animals are situated in? In order to deal with such a world, they need to exhibit certain capabilities, such as robustness against noise and failure, generalization and real-time processing. However, most traditional AI systems are lacking those capabilities. Classical approaches from *good old-fashioned artificial intelligence* (GOFAI) often view cognition as a process of world-mirroring, i.e., representing the world in form of abstract symbols in a central processing unit—a brain—that performs computations on those representations in order to control the agent. Control engineers thus try to accurately model the agent and its environment mathematically and then hard-code or hand-tune the control strategy that operates in a sense-think-act architecture, often employing if-then rules. Why is this approach flawed?

First of all, modeling can become a difficult task when artificial systems become more and more complex. Matters are enormously complicated if these are compliant systems, which are notoriously hard to model and analyze. In light of the current trend of "soft robotics" this truly becomes obstructive and calls for novel approaches [see e.g., Pfeifer et al., 2012].

Moreover, modeling must also include the agent's world. This might seem a reasonable task in artificially controlled environments such as factory facilities, but the situation is different in environments humans live in. Those are infinitely rich in features that dynamically vary on many spatial and temporal scales. It would require huge computational capacities to fully model those features and it would be a horrendous task to persistently update the model. No biological agent has such capacities and in fact, does not need them, but instead, needs to cope with some heavy amount of uncertainty, in combination with attentional mechanisms to filter out relevant information. But not only the richness of the real world challenges intelligent behavior. Additionally, the information an agent receives about itself and the world is limited. It depends on the type, number and distribution of its sensors. Those being physical devices, sensory signals are always subject to random fluctuations—noise that contaminates received information and might occlude relevant parts of it. Thus, artificial agents must possess *robustness* against noise and uncertainty in order to survive in the real world.

Furthermore, in the real world, no two situations are ever the same. Hence, the agent needs to be able to *generalize* from what it already knows to novel situations. This means it has to adapt its plans and movements to the new conditions, in order to stay flexible and cope with a variety of problems. Obviously, no if-then rule-based control strategy can provide this generalization ability in the real world, where the amount of possible situations is inconceivable.

Pfeifer and Scheier [1999] have summarized more fundamental problems that underly the classical approach. For instance, modeling the world always implies the problem of keeping the model in tune with the environment, which is dynamically changing. Additionally, symbolic representations need to be related to the real world. Symbols have meaning only to the designer or user of an artificial system. However, if it is to autonomously engage the real world, the meaning of the symbols must be grounded in the agent's interaction with the world. To solve those problems, they argue, an embodied agent must be *situated*, i.e., it must be able to perceive the world on its own by interacting with it, without a human interpreter in the loop.

Therefore, rather than detailed symbolic modeling by the human designer, the autonomous interaction with the world, involving active exploration of own capabilities and taking into account the sensorimotor processes inherent to those interactions, seems to be the key to artificial agents that fulfill the real-world requirements mentioned in the beginning.

1.2 The *pragmatic turn* and sensorimotor contingencies

Setting the focus of explanation on the interaction between an agent and its environment and explicitly taking into account the physical properties of its body seems to be an emerging key concept in all disciplines throughout the area of cognitive science. Moreover, in those interaction processes, the role of *action* for perception and cognition was revealed by many researchers as being of major importance. As Engel [2010] put it, cognitive science witnesses "a *pragmatic turn* away from the traditional representation-centered framework toward a paradigm that focuses on understanding the intimate relation between cognition and action." Rather than the capability to create world-models in the brain, which are retrieved for processes like thinking or planning, cognition should be understood as the capacity of "enacting" the world [Varela et al., 1991], as a skillful activity involving the agent's current interaction with its environment [see also Engel et al., 2013].

1.2.1 Voices of the *pragmatic turn*

First articulated by philosophers and later picked up by psychologists, such ideas have been circulating for over a century. The next paragraphs will briefly sketch the historic development of the *pragmatic turn*.

The American pragmatist John Dewey (1859–1952) stated:

> Upon analysis, we find that we begin not with a sensory stimulus, but with a sensorimotor coordination [...] and that in a certain sense it is the movement which is primary, and the sensation which is secondary, the movement of the body, head and eye muscles determining the quality of what is experienced. In other words, the real beginning is with the act of seeing; it is looking, and not a sensation of light [Dewey, 1896, pp. 358–359].

Similarly, the French phenomenologist Merleau Ponty (1908–1961) acknowledged the active nature of perception when he wrote:

> Since all the movements of the organism are always conditioned by external influences, one can, if one wishes, readily treat behavior as an effect of the milieu. But in the same way, since all the stimulations which the organism receives have in turn been possible only by its preceding movements which have culminated in exposing the receptor organ to the external influences, one could also say that the behavior is the first cause of all the stimulations [Merleau-Ponty, 1963, p. 13].

He pleaded against a representationalist view and advocated cognition as being an active process of "world-making" instead of "world-mirroring". Besides the phenomenological considerations, implications for sensorimotor learning were analyzed. The psychologist and epistemologist Jean Piaget (1896-1980) identified a pragmatic component in child development:

> There exists a sensorimotor or practical intelligence whose functioning extends that of the mechanisms of a lower level: circular reactions, reflexes and, still more profoundly, the morphogenetic activity of the organism itself [Piaget, 1952, p. 357].

Somewhat more radical rejections of the representationalist view were articulated by Varela et al. [1991], who defined cognition as *embodied action*. They emphasized that perception consists of processes for determining possible actions and that perceiving a world, i.e., being aware of it, means knowing possibilities for different actions in the current interaction:

> Thus the overall concern of an enactive approach to perception is not to determine how some perceiver-independent world is to be recovered; it is, rather, to determine the common principles or lawful linkages between sensory and motor systems that explain how action can be perceptually guided in a perceiver-dependent world [Varela et al., 1991, p. 173].

Seeking for a compromise between the cognitivist and pragmatist frameworks (see also Section 1.3), Clark [1997] grants a functional role—traditionally believed to be immanent to the brain only—to bodily and environmental characteristics. This leads him to an extended framework in which, besides aspects of the world, also possible actions are prescribed in internal states:

> ... it may for some purposes be wise to consider the intelligent system as a spatio-temporally extended process not limited by the tenuous envelope of skin and skull. [...] the traditional divisions among perception, cognition, and action look increasingly unhelpful. With the demise of the central executive, perception and cognition look harder to distinguish in the brain. And the division between thought and action fragments once we recognize that real-world actions often play precisely the kinds of functional roles more usually associated with internal processes of cognition and computation [Clark, 1997, p. 221].

From those quotations it becomes apparent that the idea of a pragmatic nature of cognition has been around for over a century (and was constantly developed further), yet it has not found its way into the mainstream theories of cognitive science and artificial intelligence.

1.2.2 Sensorimotor contingencies (SMC)

Along similar lines of argumentation as the ones from proponents of the *pragmatic turn* in cognitive science, the sensorimotor contingency theory (SMCT) was articulated by O'Regan and Noë [2001]. They define the lawful way in which one's sensory input varies as one moves around in the world as what they call sensorimotor contingencies (SMC). In their view, it is the SMCs that distinguishes perceptions from each other (e.g., vision from hearing, or using a cup from using a pen): "the *structure of the rules* governing the sensory changes produced by various motor actions" [O'Regan and Noë, 2001, p. 941, their emphasis]. According to their theory, SMCs are the key ingredients of perception, the building blocks that constitute cognition. Importantly, perception and cognition are not simply happening in the agent's brain, but are rather emerging from the interaction process itself. The brain is rather supporting cognition by enabling exercise and mastery of SMCs. Having mastery of SMCs then means to be aware of the world, by knowing the SMCs under the current engagement with the world.

SMCT has been introduced to explain mainly phenomena in visual perception of humans. However, beyond this lower, perceptional level, SMCT can also account for higher cognitive concepts such as that of objects. Objects can be defined by the patterns of regularity in the sensory inputs as a function of motor output, that are inherent to the possible manipulations of them. To perceive an object is to correctly understand that those regularities are present while interacting with it, and furthermore to understand the capacity to move and hence elicit certain changes in the sensory system.

As the example of object concepts indicates, proponents of SMCT believe that the idea on the sensorimotor level can be generalized to action-outcome contingencies, where the notion of "action" is not restricted to bodily movements, but can encompass movement sequences or mental actions [e.g., Maye and Engel, 2012a]. The generalization could hence explain higher-level cognitive processes such as planning, reasoning and decision making, which would take SMCT a critical step further towards a unified theory of cognition. In fact, the very idea is pursued by the research project "Extending sensorimotor contingencies to cognition" (eSMCs)[1]. Three types of (extended) SMCs have been proposed by the project consortium, which comprise contingencies at different levels of complexity and time [Maye and Engel, 2012a; eSMCs, 2010]:

- *modality-related eSMCs:* This first type captures the specific changes of the sensory signal in a single modality depending on the agent's action. Examples are the different perspective distortions resulting from visual sensor movements and robot locomotion, sound pressure profile changes when the robot rotates, or the dependence of the force-feedback from the force exerted by a gripper. This is the most basic type of eSMCs which, according to SMCT [O'Regan and Noë, 2001] distinguishes the qualities of sensory experiences in the different sensory channels like "seeing", "hearing", "touching" etc. This type of contingency is addressed by the SMCT in its original formulation.
- *object-related eSMCs:* The second type concerns the effects on the sensory system that are specific for each object under consideration, and are inherently supramodal. Thus, they describe the multi-sensory impression an object leaves upon a set of actions of the agent. An example is given by the different visual and force feedback signals received when touching a sponge, a piece of cardboard, or a piece of wood. This type of eSMCs identifies the object under consideration, and exercising actions from a set of object-specific eSMCs corresponds to perceiving this object. It is one of the fundamental claims of the approach pursued here, though, that the observed relations between actions and sensory changes are sufficient for recognizing a particular object. Object-specific eSMCs are more numerous and more complex than modality-specific eSMCs.
- *intention-related eSMCs:* The third type denotes a proposed novel generalization of the concept of sensorimotor contingencies and considers the long-term correlation structure between complex actions/action sequences and the resulting outcomes or rewards, which the agent learns to predict. We propose that these intention-related contingencies capture the consequences of an agent's actions on a more general level and on extended time scales. These complex eSMCs would also include contingencies that are cognitively simulated by the agent and do not relate to factual movement. After learning these eSMCs they could be used to predict if an action will be rewarding or not, and rank alternatives. At the same time this kind of contingencies can be the basis for action plans that involve several steps to reach an overall goal. Another consequence is that in this way an-

[1] EU-FP7, IST-270212, http://esmcs.eu

ticipation and anticipatory behavior, and the sense of agency could be grounded in eSMCs [eSMCs, 2010].

Extending SMCT has far-reaching implications in several areas of cognitive science. The present thesis addresses the role of SMCT (and its extensions) in artificial intelligence. How does SMCT help with building artifacts that fulfill the requirements posed by the real world? What implications does it have for perception and learning? But before addressing those issues, in order to understand how this approach is special, the following section will illustrate how artificial agents have been treated before in the quest to artificial intelligence.

1.3 SMCs in the landscape of cognitive paradigms

Where does SMCT find itself in the landscape of cognitive and perceptual theories? How is it different from other concepts and what does it share with them? This section tries to answer these questions by giving an overview of the different paradigms that cognitive science and artificial intelligence were mainly concerned with during the last century and relating them to SMCT.

1.3.1 Functionalism

In the 1950s and 1960s, the main target of explanation were aspects of higher-level human intelligence such as abstract reasoning, problem-solving, knowledge representation and planning. Those aspects, it was claimed, can be understood by modeling mental states in an abstract symbolic way in which they can be manipulated by algorithmic rules. Importantly, mental states can be identified solely with their functional role, on a level of abstraction that is independent of implementation and execution, i.e., independent of the hardware it is implemented on and of the environment it is executed in. Due to this reduction of cognitive processes to functional computation over symbols, this paradigm is known as functionalism [Putnam, 1975] (also: cognitivism [Fodor, 1975], or computationalism [Horst, 2011]). The body plays a secondary role and is simply something that is being controlled. The main emphasis is on the brain that has perceptions, beliefs, thoughts, feelings and the capability to act, as long as it is sustained by the body in the right way, e.g., feeding it with sensory information, or providing it with actuators.

The idea that cognition can be fully attributed to the isolated brain as a device performing symbolic computations [Newell and Simon, 1976] was appealing, especially in light of the upcoming personal computer, which suddenly provided a universal tool to equip artificial devices with seemingly any kind of human-level intelligence. Indeed, this kind of *good old-fashioned artificial intelligence* (GOFAI) [Haugeland, 1985] proved to be very successful in formal domains such as chess-playing, text-search or data-mining (think of search engines, expert systems or computer games, for instance). Even in the area of robotics, which involves physical interfaces to the real world, this paradigm works extremely well for applications such as manufacturing, where the plant is precisely known.

For applications in less predictable environments, this approach experienced some serious problems (more on that below), but traditional functionalism was also attacked by computational linguistics—one of its own flagships. Lakoff and Johnsen [1980] emphasized the deep entanglement between bodily experience (through interaction with the world) and the meaning and articulation of abstract concepts that is reflected in the human language.

This functionalist view on cognition is a far way from SMCT, which is of anti-representationalist nature (at least regarding the symbolic form of representations) and puts the interaction with the environment in the focus. SMCT rather speaks of perceptions as continuous, emergent patterns in the sensory and motor systems that are present during continuous interaction of the agent with its environment.

1.3.2 Connectionism

The connectionist approach [Feldman and Ballard, 1982; McClelland et al., 1986] comes closer to that view, as it acknowledges the emergent nature of mental behavior. It regards mental phenomena as resulting from the interplay of distributed and parallel computations, performed in networks of interconnected simple processing units (the most prominent example being artificial neural networks). After having been abandoned in the late 1960s, when Minsky and Papert [1987] demonstrated the limits of single-layered Perceptrons and their inability to handle even simple functions like XOR, the connectionists approaches became very popular among scientists during the 1980s, when the power of neural networks was recognized through algorithms like Backpropagation. Although *emergence* is acknowledged for the first time, connectionism still limits its cognition to processes within the brain and ignores the properties of body and environment, or their interplay with neural processes.

1.3.3 Embodiment, situated and behavior-based robotics

The largest criticism on the functionalist paradigm came from the area of behavior-based robotics and artificial intelligence, where the sense-think-act architecture appeared to be an obstacle for even simple real-world interactions [Brooks, 1986]. Here, information about the environment is sampled through physical sensors, noisy and incomplete and actions are mediated by physical devices with imprecise outcomes and dynamics that impose tight temporal constraints [Pfeifer and Scheier, 1999]. Moreover, it was shown that already with minimal or even without computational processing performed by a brain or controller, artificial devices can achieve behaviors that one would call cognitive, only by simple neural structures or by clever mechanical design of the body. The most illustrating examples are the Braitenberg-vehicles [Braitenberg, 1986], the tortoises by Walter [1953], or the passive dynamic walker [McGeer, 1990].

One key insight was that behavior cannot be sub-divided into components that play functionally distinctive roles. Rather, behavior *emerges* as a result from the recurrent and nonlinear interplay between brain, body and environment [Steels, 1993; Pfeifer and Scheier, 1999]. Instead of a serial chain of computational routines, in order to explain this emergent nature, the usage of Dynamical Systems Theory was suggested to be a more appropriate modeling framework, where no privileged status was attributed to anything like a "central processing unit" [Beer, 1995]. Bodily and environmental characteristics play as central a role in a rather fluid and flexible sensorimotor coupling that leads to cognition and intelligence.

SMCT comes very close to this view, in putting the sensorimotor coupling in the main focus. SMCT and its extension go further by claiming that perception and cognition are *defined* by this coupling—these contingencies—and that their specific structure is what causes percepts of concepts like objects or colors.

Undoubtedly, early attempts of the embodied, situated and behavior-based approach could explain intelligence merely on a basic level, far away from cognitive capacities like abstract reasoning or problem solving. However, it was argued that in the history of evolution the achievement of basic insect-level behavior took exponentially more time than the remaining steps to human-level cognitive capabilities. Similarly, in the study of cognition, the hard part is to first understand seemingly simple, yet robust and adaptive behavior, after which understanding of higher-level intelligence will follow more easily [Brooks, 1990, 1991]. With similar arguments, Beer [2008] proposed to study the emergent dynamics in the interplay between brain, body and environment first in simple model systems, in order to develop the tools necessary to analyze complex animal and human systems in future.

1.3.4 Embodied cognition, enaction and extended mind

The embodiment movement had reached also researchers beyond robotics (e.g., psychology, linguistics, neuroscience, philosophy, general artificial intelligence). Beyond agreeing on the major importance of the body and its interaction with the world for the behavior of an agent, they soon collected evidence that also higher-level cognition is actually grounded in this interaction and bodily properties.

In the enactivist approach, the separation between body and brain is discarded altogether and the body is seen as a whole [Varela et al., 1991; Di Paolo et al., 2010]. This further lead the proponents to the body being constitutive for concepts like agency, autonomy and autopoiesis, where its identity is closely tied to autonomous processes of self-construction. An important aspect is the active role of the agent: in order to make sense of the world it must actively engage it. Through this active engagement or *enaction* only, cognition is emerging.

The active touch of "world-making" is fully in line with SMCT which similarly claims that stable perceptions of the world do not occur through matching of stimuli with internal world models, but can only form by actively interacting with it ("the world as an outside memory" [O'Regan and Noë, 2001, p. 1019]).

The traditional functionalist paradigm was also further developed to account for the legitimate criticism from proponents of embodiment. The extended functionalism (also: extended mind theory) [Clark and Chalmers, 1998; Wheeler, 2011], still abides by the view that cognition can be seen as functional computation, but it is extended to enclose not only the brain as processing device, but also other physical media such as the body and even the environment. Computation is "outsourced" from the brain to body parts, tools and objects that are involved in the sensorimotor processing stream. Those bodily and environmental features form an extension of the implementational resources that preserve cognitive computations. In that sense, embodiment acts as pre-processing of informational signals through an appropriate sensory apparatus (cf. *information-self-structuring* [Pfeifer et al., 2007]) The extended mind theory thus shares ideas with *morphological computation* that demonstrates how a physical system such as the body of an agent can perform computations, provided that it has a moderate degree of compliance or softness [Pfeifer and Bongard, 2007; Hauser et al., 2012; Pfeifer et al., 2012].

1.4 Thesis overview

Can the action-centered, pragmatic view lead to more robust perception, learning and control frameworks for artificial agents than classical GOFAI approaches provide? What requirements must be fulfilled in order to apply the sensorimotor contingency approach to robots? How does SMCT integrate into existing concepts such as that of internal models, body schema, or reinforcement learning?

In an attempt to answer those and related questions, the collection of papers presented in this thesis investigates the applicability of the theoretical concepts of (extended) sensorimotor contingencies to artificial agents from several different viewpoints. After providing background information about related work and concepts relevant to our approaches in Chapter 2, the main results and achievements are summarized in the subsequent chapters, while leaving the details of methodology to the attached papers. Implications of an information-theoretic view on SMCT are investigated in Chapter 3. Here, we investigate the applicability of information measures to sensorimotor recordings of a robot and show that such measures can successfully extract modality- and object-related SMCs, leading to a basic body schema of the robot. This is followed by a discussion on aspects of SMCT in the context of reinforcement learning in Chapter 4. We show

how a probabilistic Markov model can capture the SMCs of a robot and, by providing feedback in form of rewards, intention-related SMCs can be learned and utilized to maximize an internal value. Chapter 5 gives some thoughts to a modeling perspective of SMCT, taking human-level sensorimotor coordination as a model for artificial cognition. After presenting work on learning forward and inverse kinematic models for hand-eye coordination, we present a model of human visual attention that might serve as a basis to model robot visual exploration. Finally, Chapter 6 closes this thesis with a discussion of the presented results and their implications on the above mentioned questions.

Chapter 2

Related Work

Before presenting the results of our own investigations in the forthcoming chapters, this chapter reviews some important concepts and related work on which our studies are based on or which show parallels to our approaches that are worth mentioning.

2.1 Related sensorimotor approaches.

The methods presented in this thesis are not the only existing implementations of SMCT, although models that faithfully implement SMCT are scarce. Several other existing methods approach cognition with a sensorimotor approach, often without directly relating it to SMCT. Here we list a few and discuss their relation to SMCT.

Pfeifer and Scheier [1997] present models that use artificial neural networks connecting sensors and motors to store SMCs. The extended Braitenberg vehicles equipped with this architecture are able to learn to approach small and avoid large objects by interacting with them.

Hebbian learning is used to associate sensory and motor channels in [Bovet and Pfeifer, 2005]. However, no distinction is made between sensors and motors, i.e., all signals are equal on initialization. A form of Hebbian learning between motor and sensory signals is also used in [Marques et al., 2014] to learn reflexes from spontaneous muscular activity in a musculoskeletal leg model. By relatively simple top-down modulation of the established reflex circuits the model can then produce coordinated behavior.

The system presented by Möller and Schenck [2008] uses neural networks to learn forward and backward models of an agent's sensory changes in relation to preceding actions. The models are learned incrementally from interaction and are used to perform long-term simulations and distinguish between dead ends and corridors (without representing the two situations in the sensory domain). As will be discussed in Chapter 5, internal forward models (and their inverse counterparts) are SMC models if they are used to predict sensory states from sensory and motor inputs (or necessary motor input from desired sensory states).

Fine et al. [2007] present a simulated agent with internal dynamics evolved to enable phototaxis, which is able to adapt to large disruptive changes in its sensor morphology. Here the sensorimotor relations in individuals of one generation are established using a mutation scheme and selecting individuals according to some fitness function for the next generation. Similarly, the vast work on reinforcement learning (including the work that will be presented in Chapter 4) uses evaluation of performance in form of utility functions. Those methods can be interpreted as SMC models, for the cases which use actions and sensory states as inputs to the prediction of the

utility and, importantly, if the SMCs are utilized for action-selection as argued by Maye and Engel [2013].

Learning motor primitives (also: muscle synergies) is a way of reducing a high-dimensional motor space to a lower-dimensional control space. Looking unrelated at a first glance (missing relation to sensory changes), this paradigm shows interesting parallels to SMCT: recent developments in motor primitives research revealed the relevance of the task-space for formation of primitives [Alessandro et al., 2013]. This means the control space (action) needs to be developed with regard to the task, mediated through sensory or higher-level outcome, which can be viewed as a higher form of SMCs.

The concept of incremental learning by intrinsic motivation proved successful in the area of developmental robotics [Oudeyer et al., 2007; Oudeyer and Kaplan, 2007]. Here the goal is to implement a mechanism that evaluates the degree of novelty or surprise that a situation provides (therefore it is also called artificial curiosity [Schmidhuber, 1991]). Often this novelty measure is expressed as a reward in a reinforcement learning setting, where the reward is maximal when this measure is in an intermediate level. Baranes and Oudeyer [2013] use the intrinsic motivation to guide learning of inverse kinematics with a redundant robotic arm and of locomotion in a quadruped robot using motor synergies, both not modeling SMCs directly, but the learning methods include capturing of SMCs as a premise.

2.2 Recent approaches to study SMCs in artificial agents

While discussion about implications of SMCT in philosophy and psychology is ongoing, very few attempts have been made to define SMCs formally in order to develop models that can be used in artificial systems. Here we highlight two contributions that aim at either directly implementing SMC models for robots or deriving a formal language with which SMCT can be further modeled and tested. These works are especially important for the studies presented in the forthcoming chapters and will be recurrently referred to.

2.2.1 A markov model of SMCs

Maye and Engel [2011, 2013] proposed a computational model for SMCs that uses probability distributions of the sensorimotor space to store the relations between actions and accompanying observations. They formalized the interaction of an agent with the world as a generalized Markov decision process, where the transition probability distributions $p(s_t|s_{t-1}, a_{t-1}, ..., s_{t-h}, a_{t-h})$ reflect the SMCs (s_t and a_t are sensory states and actions at time step t, respectively). They claimed that the time horizon h of the distributions reflects the *type* of SMCs: modality-related SMCs are characterized by the immediate changes in the sensorimotor signals (small h); object-related SMCs are the relations inherent to the manipulations of objects and live on a longer time scale, thus a larger h [Maye and Engel, 2012a]. In their model, the SMCs were utilized for learning by extending the system with a value system that allows to directly use the acquired SMCs for goal-oriented action selection. Their model was demonstrated on simulated and embodied robots with a rather low-dimensional and discrete sensorimotor space (1 sensor, 1 motor) in object discrimination tasks with two object classes. The agents could successfully recognize the objects and trigger object-specific actions. Later the model was extended from the prediction of immediate events to prediction of longer interaction sequences using previously experienced SMCs [Maye and Engel, 2012b]. Results were demonstrated on a robot with a one-dimensional action space and seven-dimensional sensor set, using a utility function determining the task.

Their approach constitutes a direct straight-forward implementation of the SMC concept as stated in [O'Regan and Noë, 2001]. A strong focus of these concepts and the computational model

is the direct incorporation of action context in the perception and recognition of objects. Whether the recognition performance is actually enhanced through the action-context, however, remains to be tested. Furthermore, the conceptual claim of history-dependency and its relation to modality- and object-related SMCs needs to be substantiated by experimental evaluation.

As explained above, their approach uses conditional probability distributions of action-outcome contexts of variable length. This raises some implementational issues regarding storing and searching the experienced distributions, which can become high-dimensional and sparsely populated if action or observation spaces become larger. Finally, the question remains, how to deal with previously unseen samples. The model is not able to generalize from its experience to novel situations, but instead selects a random action in those cases. In our work presented in [Hoffmann et al., 2012], appendix D, we picked up on this approach to answer some of these open questions, and in [Schmidt et al., 2014], appendix E, we extended the model to tackle some of its shortcomings.

2.2.2 A dynamical systems account for SMCs

The sensorimotor contingencies theory as articulated by O'Regan and Noë [2001] lacks a clear definition of how SMCs can be formalized. Buhrmann et al. [2013] set out to enable empirical modeling and testing of SMCT and arrived at an operational definition of SMCs in dynamical systems terms (see also [Di Paolo et al., 2014]). This provides a formal language that can be used to address SMCT and facilitate the conceptual clarification. They propose four distinguished notions on different levels of cognition. First, they define the *sensorimotor environment*, as what encloses all the relations between motor and sensory patterns of all possible interactions of agent and environment. These are all objectively possible relations of the coupled physical system and are completely independent of the internal dynamics of the agent—an open-loop system. When closing the loop and considering the agent's internal dynamics for motor activity, the possible sensorimotor trajectories are constrained to a set of relations they call the *sensorimotor habitat*. This subspace of what can actually be achieved by the agent (within all physically possible trajectories) might be much lower-dimensional. The next level includes the notion of a task domain and denotes relations from the SM habitat that are functionally relevant for the task performance. These relations are referred to as *sensorimotor coordination*. They are again a subset of the lower level, but are restricted to those parts that have functional significance for the task. Finally, many SM coordinations might achieve the same task or might have to be activated together, in order to achieve a task. A normative element arises which evaluates, which SM coordinations are more precise, fast, or efficient in achieving a task. By introducing this normative use of SM coordinations, which they refer to as *sensorimotor strategy*, Buhrmann et al. [2013] made a connection to the vague notions of "mastery", "skillful activity" or "laws of SMCs" many times mentioned in [O'Regan and Noë, 2001] and that are making SMCT so difficult to formalize.

By applying those categories of SMCs to a minimal agent model, the authors demonstrate that it is now possible to qualitatively describe the "laws of SMCs" as topological properties of the sensorimotor space, such as symmetries, smoothness, attractor regions and so on. Furthermore can the laws of one level be constraining those of other levels. Whereas their work does not directly propose a framework to *learn* the different levels of sensorimotor relations, they argue that the knowledge of sensorimotor co-variations does not necessarily need to be stored explicitly in order to achieve a task. Instead, it can be sufficient for the agent to fixate the task-relevant SMCs in its internal dynamics.

The classification of SMC types here is different from our working definitions proposed in Section 1.2.2 (modality-, object- and intention-related SMCs). Starting at the SM environment, the state space of possibly occurring sensorimotor trajectories is narrowed down in each level as internal dynamics or intentions are introduced. At the same time they account for a potentially

increasing dimensionality, by including task-relevant information (performance measures) or parameters to select between sets of trajectories. Buhrmann's definition does not account for concepts like modalities or objects. Both live as specific features of sensorimotor trajectories in the SM habitat, and enter the SM coordination/strategy domain as soon as they become relevant for a certain task. Intention-related SMCs find their counterpart in the trajectories affecting additional dimensions that measure task performance and are introduced in the SM coordination/strategy. We discuss those categories of SMCs and their relations further in Section 6.2.

2.3 Situated perspective and the brain-body-environment system

Throughout our work, artificial and biological agents are treated as embodied entities, situated in their specific ecological niche. In this perspective, the interaction of an agent with its environment is typically modeled as an interplay between three components: the body, the environment and the brain or controller [e.g., Pfeifer and Bongard, 2007; Beer, 2008; Polani et al., 2007]. These components are determining the SMCs an embodied agent can experience, and thus contributing in different ways to the emergence of cognition. First, the agent's body has a certain morphology, e.g., with limbs of specific shape and orientation, having certain materials on different parts, etc. The body provides sensors of different kinds that are distributed in a particular way, enabling the agent to receive information about its current state and that of its environment. And the body provides certain means of actuation to alter its configuration or manipulate its environment. These morphological properties affect the SMCs of this agent. For instance, having both eyes at the front of our heads provides a certain relation between the motor signals inducing eye movements and the sensory signals from both eyes. This relation would be completely different if one eye would be e.g., on the right hand. And this different relation would also change the potential for cognitive capabilities (e.g., what happens to depth perception in this case?). Second, the body is controlled by a brain or controller that follows certain rules by linking sensory stimulations and internal states to motor actions. If these rules change, the behavior is different and the experienced SMCs are altered. Finally, the environment the agent lives in also sets certain constraints on the SMCs. Light conditions, properties of objects that are around, the behavior of other agents in the environment: all this influences how sensory patterns are structured and which actions are possible during interaction. These characteristics of brain, body and environment are determining what Buhrmann et al. [2013] call sensorimotor environment: All functional relations between sensory and motor signals that are possible given the specific embodiment, internal dynamics and external environment of the agent.

2.4 Incremental learning from interaction

A main theme throughout this thesis is the gathering of sensorimotor knowledge by artificial agents through autonomous exploration of their capabilities. Rather than providing external knowledge e.g., through mathematical modeling of motor trajectories, an agent needs to extract relevant sensorimotor information on its own from its interaction experience. This way, the agent's behavior is grounded in its embodiment and situatedness, which prevents a frame-of-reference fallacy, in which learning could be biased or misled by the experimenter's view of the problem [Pfeifer and Bongard, 2007, p. 72ff.].

In order to solve a task without prior knowledge, a good strategy seems to be to start with random exploration of the sensorimotor environment or habitat, i.e., to move the agent's actuators

randomly, thereby self-generating sensorimotor streams from which one can capture the relations between the actions and accompanying changes in the sensory system as well as changes in the reward system. Such a process has been called "motor babbling" in the artificial intelligence community [Demiris and Dearden, 2005; Der and Martius, 2006] and is inspired by the seemingly random movements that are performed by humans in their early development [Meltzoff and Moore, 1997; Bernstein, 1967].

The motor babbling approach has been used in several approaches before. In robotics, such a strategy is typically either combined with learning of internal models of the agent which can be used to predict the outcome of certain movements or to generate actions that lead to desired sensory changes (see [Nguyen-Tuong and Peters, 2011] for an overview). Alternatively, when integrated with a higher-level feedback of what is desired and what is to be avoided—a reward system—the agent can use the acquired SMCs to learn desired behaviors in a reinforcement learning fashion [Sutton and Barto, 1998; Kober and Peters, 2010; Murao et al., 2001].

We will make use of those concepts throughout this thesis. In Chapter 3 we use the motor babbling method to capture modality-related SMCs (or SM environment) of the agent, from which we derive a primitive body schema. In Chapter 4 the object-related and intention-related SMCs are learned (SM strategy) and in Chapter 5 adaptation mechanisms exploit this strategy, narrowing the SM environment towards its SM coordination.

Chapter 3

Information-Theoretic Account for SMCs

3.1 Introduction

In this chapter, the potential of information-theoretic measures for learning of SMCs in artificial agents is investigated. The application of such measures is twofold: as generic analytic tools for inspecting the interaction processes in complex systems and, more interestingly for our purposes, as unsupervised statistical learning methods for autonomous agents that enable them to extract useful information about themselves and their environment from their sensorimotor streams.

In this perspective, information theory is used to capture SMCs and identify them as informational structures that are specific to an agent's embodiment and specific to its interaction with the world. Different to other methods such as those reviewed in section 2.2, here the SMCs are not directly utilized for solving certain tasks. Rather, it is shown how such an information structure helps the agent to acquire a basic understanding of its body and of the nature of its sensors and actuators by generating a primitive body schema. Moreover, it is shown that those measures provide the agent with means to distinguish objects in the environment by their specific information "fingerprints" the interaction with them induces in the sensorimotor stream. Finally, this chapter gives a prospect of how information-theoretic measures could provide the agent with a method for learning coordinated, yet task-independent behavior.

The work presented in this chapter is summarized in [Schmidt et al., 2013, 2012], appendix A and appendix B, and in [Nakajima et al., 2015], appendix C.

3.1.1 Information transfer in embodied agents

Let us consider how a naive embodied agent perceives the world as it interacts with it. The agent receives a high-dimensional stream of sensorimotor signals which it has to make sense of, in order to enact the world in a meaningful way (non-random, coordinated, task-oriented). Given no prior knowledge about itself and the world, the best way to start, we argue, is to extract the statistical properties and correlations between these sensory and motor signals. These correspond to the SMCs, the laws describing how signals change in a systematic way when the agent moves around.

We identify and quantify these relations using tools from information theory, because these provide us with a generic currency—the Shannon information—which is universal in the sense

that it is not specific to the type or magnitude of signals we encounter. In contrast to simpler correlation measures such as covariance or Pearson's correlation coefficient, information measures capture non-linear relations and are usually extendable to multivariate cases.

Specifically, we use transfer entropy [Schreiber, 2000], as it provides a nonlinear and directed correlation measure between two signals. Transfer entropy is a specific form of conditional mutual information (see e.g., [Cover and Thomas, 2006]) that provides directionality and uniqueness by conditioning on past states of the target variable (in contrast to time-delayed mutual information, which also provides directionality, but retains some shared information present in both signals) [Williams and Beer, 2011]. Transfer entropy can be viewed as measuring the predictability that one signal adds to the transition of another signal. If the transition of one signal to a new state can be predicted to a certain extend by its own history, the transfer entropy quantifies how much this prediction can be improved by looking at the second signal. Thus it represents the unique information that is transferred from one signal to another. As such, transfer entropy qualifies as a quantification measure for SMCs, when applied to the sensorimotor signals of an agent, especially when measuring the information transfer from motor to sensory signals. Transfer entropy has been widely used to study information transfers in the brain (see [Wibral et al., 2014] for a recent review) as well as for analyzing behaviors in artificial agents [Lungarella and Sporns, 2006; Williams and Beer, 2010; Nakajima et al., 2011]. It is noted, however, that other information measures might be suitable as well to extract SMCs, if one wants to emphasize other characteristics of the relations than uniqueness and directedness (see e.g., [Lungarella et al., 2005; Hlavackovaschindler et al., 2007; Ay and Polani, 2008]).

To test transfer entropy as a measure for SMCs, we applied it to the sensorimotor recordings of an artificial agent. The agent we used in a case study was the Puppy robot, a quadruped machine with four motors at its hip joints and legs that consisted of two segments connected by a passive knee joint and a spring that pulled the joints to a resting angle. The robot had a set of 18 sensors of four different modalities. We provided three different sets of motor signals in an open-loop manner, two gait-patterns (bounding and turning) and one set of independent random signals. The details can be found in [Schmidt et al., 2013], appendix A.

3.2 Impact of the controller on SMCs

The transfer entropy values obtained from the three experimental conditions are shown in the matrices in Figure 3.1. It can be seen that the overall information transfers were much higher in the gaits (Figure 3.1 a,b) than in the random setting (Figure 3.1 c, note the differently scaled color bars). Furthermore, there were systematic patterns in the three control settings, that showed high "inter-leg" dependencies in the gaits, but only "intra-leg" dependencies in the random condition (e.g., high information transfers from motors to hip angles and knee angles of the same leg). This was because of the coordinated movement patterns that the gaits induced. Their motor signals were highly synchronized between different legs and were moving in an oscillatory way, therefore providing a lot of information about each other (and about the sensors attached to other legs as well). In the random condition, the legs moved independently and therefore provided only little information about each other. However, the signals belonging to sensors or motors within one leg provided some information about each other due to the strong mechanical couplings. The highest being the motors influencing the angles of the hip joints they were attached to and the knee joints and foot pressure sensors in the same leg, respectively. Furthermore, we note the different informational structure between the bounding and turning gait. The legs whose motor signals were synchronized or in phase were different ones in both gaits. Thus, the information transfer between certain legs could be high in one gait, but low in the other one.

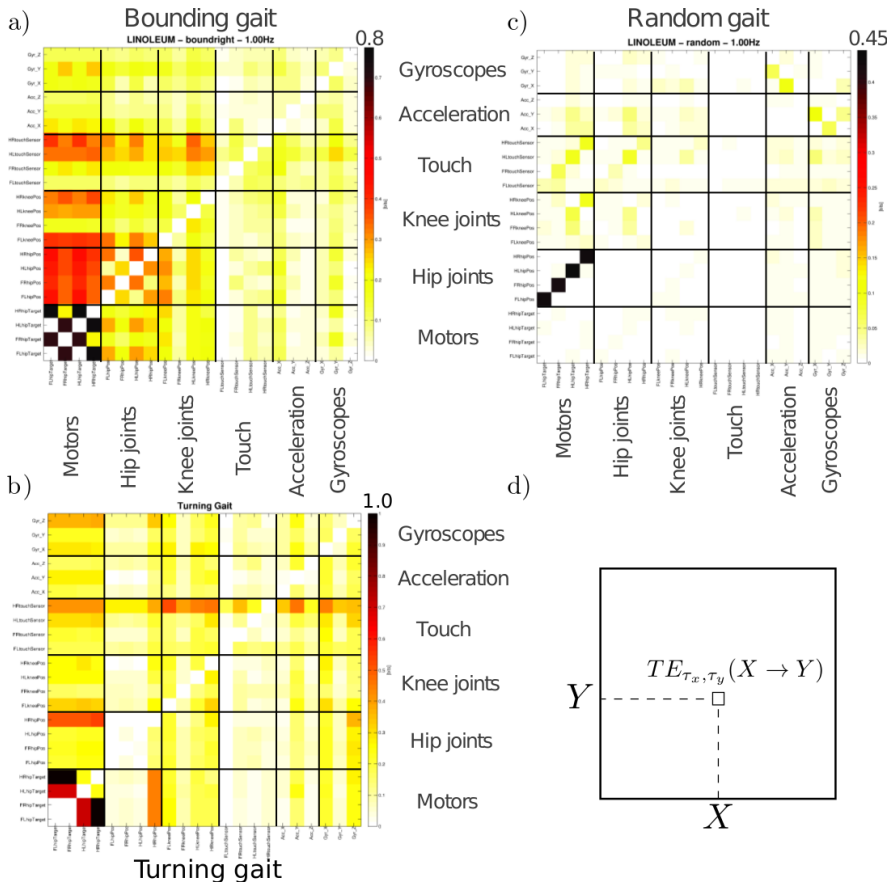


Figure 3.1: **Transfer entropy among the sensorimotor signals in the Puppy robot. (a-c)** The matrices show the information transfers using two evolved gaits (a,b) and one random set of motor commands (c). **(d)** Each cell depicts the transfer entropy from the signal in the column to the signal in the row. The signals are grouped according to their modality and are ordered according to the leg where the respective sensor (or motor) is located (front left, front right, hind left, hind right) or according to the axis in 3D space (for the acceleration and gyroscope signals). Figure taken from [Schmidt et al., 2012].

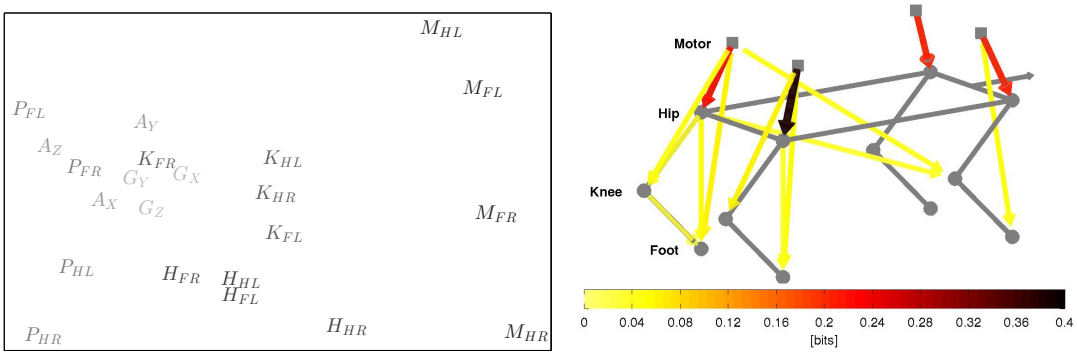
One major observation of these results is that the information transfer depends on the type of control signal used. The "brain"-part of the brain-body-environment system is influencing the information structure revealed by transfer entropy, a property inherent to SMCs. Moreover, coordinated control signals seem to induce high information transfers between the sensorimotor signals. This seems to suggest that, in order to achieve coordinated movements, an agent should try to strengthen its SMCs by appropriate modification of its internal dynamics.

3.3 Learning the SM environment from motor babbling

The independent random motor commands seem to reveal some basic SMCs inherent to the embodiment of the agent, given the world it lives in. In theory, using a random motor command and assuming the robot visits all possible sensorimotor trajectories during the experiment, can be

seen as "marginalizing out" the brain part of the system, leaving the SMCs that are specific for the properties of the body and environment of the robot. In Buhrmann's terms (see Section 2.2.2), this corresponds to the SM environment. Of course, in practical experiments it is impossible to visit all sensorimotor trajectories that a given robot could visit (at least for fairly complex ones), but this example shows that the transfer entropy is in principle capable of revealing an approximation of an agent's SM environment.

When taking the differences between individual columns and rows of the obtained information matrices in Figure 3.1 as a metric to project the sensory and motor channels into a 2D space using multidimensional scaling, a *sensoritopic map* can be derived, as it is shown in Figure 3.2a. It reveals that the different modalities group nicely in the information domain, concluding that the transfer entropy provides a basis for capturing the structuring of the sensorimotor space into several modalities, corresponding to modality-related SMCs. These informational similarity-based modality-related SMCs, together with the topological ordering revealed from the SM environment result in what can be called a rudimentary body schema.



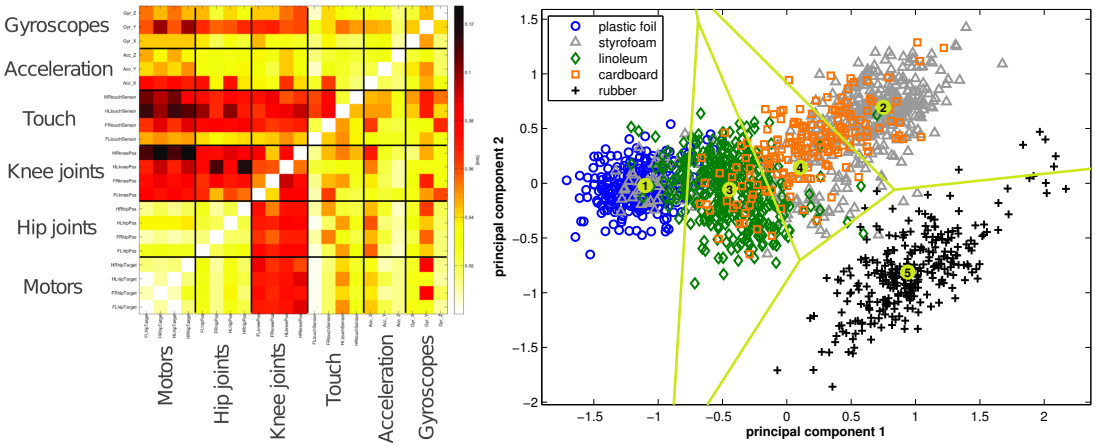
(a) Sensoritopic map

(b) Body schema

Figure 3.2: **From SMCs to a body schema. (a)** Projection of the sensors and motors into 2D space using multidimensional scaling based on a similarity measure of information transfers. The abbreviations mean the following: motor signals M, hip angles H, knee angles K, touch pressure P, acceleration A and gyroscope G. The subscripts are: front left/right leg FL/FR, hind left/right leg HL/HR. Figure taken from [Schmidt et al., 2013]. **(b)** Information transfer values of the random control condition mapped on the topology of the robot. Thickness and color of arrows indicates the strength of information transfer among channels. Figure taken from [Schmidt et al., 2012].

3.4 Impact of the environment on SMCs

Additionally to the control scheme, the environmental conditions were varied systematically. The robot ran on five different ground materials, which differed in friction properties and surface structure (see [Schmidt et al., 2013], appendix A for details). The analysis showed varying patterns of information transfer when the robot ran on different terrains (Figure 3.3). The matrix in Figure 3.3a reveals which SMCs varied the most across terrains. On an abstract level, the different terrains can be viewed as being different objects the robot was interacting with. The relations or SMCs that were specific to certain terrains could thus be identified as object-related SMCs. They were only present during interaction with the specific terrain and vanished as the agent interacted with another terrain (whose SMCs were then activated instead). Figure 3.3b shows that



(a) Std of transfer entropy across terrains

(b) PCA projection of transfer entropy

Figure 3.3: **Object discrimination based on information transfers. (a)** The standard deviation of the transfer entropy matrices across different terrain conditions. **(b)** Terrain clustering based on the transfer entropy matrices of all experiments. For visualization purposes, the matrices are projected onto the first two principal components. Figure taken from [Schmidt et al., 2012].

the terrains could indeed be discriminated using the information transfer among sensorimotor channels.

3.5 Method extensions: symbolic, delayed and time-local information transfers

The analyses described above used the transfer entropy measure in its original form [Schreiber, 2000]. In this section, some improvements of this measure are introduced, which solve problems that occur when dealing with real-world data of high-dimensional oscillating systems. A detailed investigation of these improvements is summarized in [Nakajima et al., 2015], appendix C. Using these extensions of transfer entropy, (1) the required preprocessing of raw sensorimotor data is minimized, (2) delayed couplings are captured, (3) the spatiotemporal resolution of multivariate information transfers is increased and (4) the timing of information transfer can be localized, which is especially interesting for oscillating systems such as the periodic gait patterns in the Puppy robot.

In our study [Nakajima et al., 2015], appendix C, the extended transfer entropy measures were applied to the recordings of the soft octopus arm robot. This octopus-inspired robot had a silicon rubber arm with two embedded cables for actuation by two pulling motors and was operated in a water tank. In two different conditions the motors were programmed to pull the cables (1) randomly and (2) in turns, leading to oscillatory arm movements. In a third condition an obstacle was placed in the robot's vicinity during oscillatory movements. Signals from force sensors measuring the cables' tensions were recorded and the positions of the arm segments were tracked in camera images.

3.5.1 Handling real-world data by permutation entropy

information-theoretic measures are functionals of probability distributions. Estimating those distributions from the obtained real-world time series, which are usually discretized and noise-contaminated, is an ongoing research topic. Many simple estimators introduce biases or require careful fine-tuning of their parameters. One possible treatment of these issues is to apply probability distribution estimators that are based on permutation partitioning of the data points [Bandt and Pompe, 2002]. The resulting uncertainty measure *permutation entropy* quantifies the uncertainty of the local orderings of data values, instead of the uncertainty of the values themselves. This procedure can be applied directly to raw data and does not require further model assumptions such as the range of the data values present in the time series, or the precision of their discretization. information-theoretic measures based on permutation have been proven to be easy to implement, fast to compute, robust against noise, as well as applicable to non-stationary and transient time series data. Since this approach does not require any prior knowledge about the nature of the time series, it is especially suitable for unsupervised extraction of SMCs in complex artificial agents. The extended information transfer measures being introduced in the following sections are based on this permutation type entropy.

3.5.2 Capturing delayed couplings

The transfer entropy measure fails at detecting delayed information transfer, which is why the time horizon of a detected information transfer is always unspecified. By using a specific conditioning on the past states of the source and target variables in question, this drawback can be overcome [Pompe and Runge, 2011]. The derived measure *momentary information transfer* thus has a better resolution for point-to-point information transfers in the temporal domain. We used the permutation equivalence of this measure, the *momentary sorting information transfer* (MSIT), in [Nakajima et al., 2015], appendix C and demonstrated its capability of detecting delayed SMCs in the soft octopus arm robot. Figure 3.4a shows the information transfer from the force sensors at the cables to the arm's tracked positions i for several time lags $\tau = 1 \dots 20$. The results show that the information transfer to the base of the arm was almost instantaneous, whereas that to the tip took about 7 time steps (~ 0.35 s).

3.5.3 Increasing resolution in multivariate time series

So far, in our analyses in the Puppy robot, only bivariate information transfers were considered. In this case, we used a conditioning on past states of the target variable in order to exclude shared information, which resulted in a high-resolution measure. When accounting for multivariate time series, we have to take particular care of other possible information contributors apart from the information source under consideration, in order to keep the high resolution. This requires some extra conditioning to "filter out" the effect of these other information sources, which might otherwise occlude the effect of the source in question. When estimating SMCs in high-dimensional sensor sets with tight structural couplings, this is especially relevant. Even more so, when dealing with compliant or soft robots with highly interconnected topologies. However, in order to perform further conditioning on neighboring information sources, a certain knowledge about the topological ordering in the platform is required (i.e., we need to know which sensors are adjacent). Furthermore, it is assumed that the degree of information transfer between components depends on their relative spatial and temporal distance, which can be safely assumed for most physical systems. Figure 3.4b visualizes how the momentary sorting information transfer with extra conditioning works. In [Nakajima et al., 2015], appendix C, we applied it successfully to enhance detectability of information transfers within the arm of the soft octopus robot.

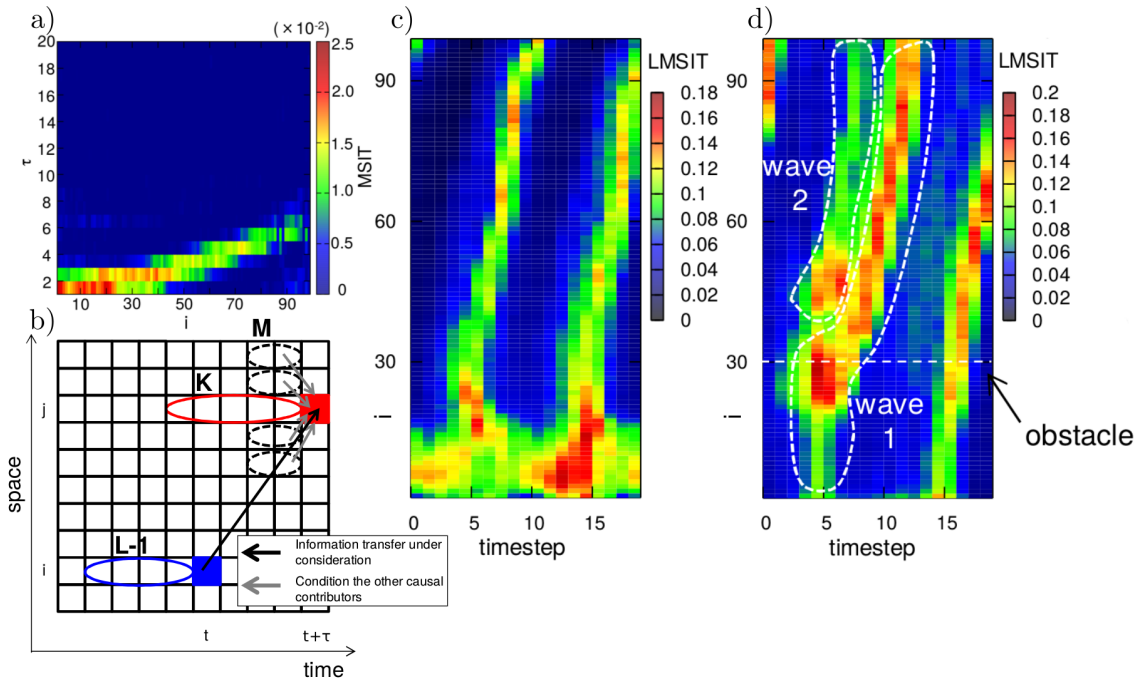


Figure 3.4: **Extended transfer entropy methods in the octopus arm robot. (a)** Momentary Sorting Information Transfer (MSIT) from force sensors to positions of the arm segments i for several time lags τ during random movement. **(b)** Schematic expression of the extra conditioning of MSIT on neighboring segments of the arm. The black arrow shows the information transfer under consideration, the gray arrow shows the influences of other sources. **(c)** Local Momentary Sorting Information Transfer (LMSIT) from arm segments i to $i+1$ within one oscillation (20 ms) of the oscillatory movement. **(d)** The same LMSIT in the condition with the obstacle in the water tank. The horizontal white dashed line shows the contact point of the arm with the obstacle, the two white dashed regions indicate two "information propagation waves". Figure adapted from [Nakajima et al., 2015].

3.5.4 Time-local information transfer

The last extension makes it possible for our information measure to characterize a local information transfer profile within the agent's spatiotemporal dynamics. The transfer entropy as well as the momentary sorting information transfer introduced above are, due to their statistical nature, averages across the whole time series. The measures provide no further specification about the time at which the information is transferred, i.e., from which point in time in the source variable it is transferred to which point in time in the target variable. However, this timing might be interesting and is very specific to the material properties of the system. For instance, an impact on one end of a metal rod can be measured almost instantaneously on the other end, whereas in a rubber rod it takes some time to propagate the impact through the material. Especially in oscillatory systems such as the Puppy robot, which runs using repetitive motor patterns (gaits), it is interesting to observe when the information transfer from one signal to another occurs within the locomotion periods. This local timing is a more detailed characteristic of the SMCs which is desirable to extract. In [Nakajima et al., 2015], appendix C, we introduced the *local momentary sorting information transfer* (LMSIT) that fulfills this purpose. In the study it was used to track down the information

propagation within the soft octopus arm robot for observatory purposes, but it provides a suitable method for learning time-specific characteristics of SMCs in general in oscillating systems. Figure 3.4c,d illustrate the timing of information transfer within an oscillation of the oscillating octopus arm movement. The LMSIT from each tracked arm segment i to its neighboring segment towards the tip $i+1$ is shown. The image reveals the precise timing and speed of the information transfers during an downward movement of the arm ($t = 5$) and subsequent upward movement ($t = 15$) as two "information propagation waves". When an obstacle was put underneath the arm (Figure 3.4d), the impact of the arm on the obstacle manifested itself as a split of the first propagation wave into two separate waves. One with the normal speed that could be attributed to the robot's body properties (wave 1 in Figure 3.4d) and one quicker wave that could be attributed to the impact on the obstacle, i.e., the interaction with the environment (wave 2 in Figure 3.4d).

3.6 Summary

In this chapter, techniques from information theory were introduced as suitable unsupervised learning tools for capturing SMCs in artificial agents. The results presented here show that transfer entropy (and related information-theoretic measures) in combination with motor babbling provide a learning strategy that does not require any knowledge about the agent or its environment. With this, a primitive body schema of the agent can be created by grouping sensory channels into modalities according to their information transfer profiles. The information transfer structure can furthermore provide a sense of which channels are sensitive to own body configurations and which ones provide information about the environment, deriving a degree of exteroceptiveness and proprioceptiveness of sensors.

The transfer entropy exhibits some characteristics of SMCs, such as the sensitivity to changes in individual components of the brain-body-environment system. We demonstrated this for the brain-part (controller) and for environmental features, revealing that the measure is not only capable of capturing the SM environment and modality-related SMCs, but moreover allows to learn object characteristics that we identify as object-related SMCs.

Furthermore, our data suggests that information transfers could be a drive to guide learning of coordinated behavior by adjusting the control dynamics to maximize information transfers.

Finally, in this chapter, a number of extensions for the transfer entropy method were introduced, which increase practicability for real-world agents and allow for extraction of more characteristic features inherent to SMCs, such as precise temporal locality or filtering from a bunch of informational sources.

Chapter 4

SMCs in Reinforcement Learning

4.1 Introduction

This chapter discusses the role of SMCs in learning of control policies for solving predefined tasks. Again, an important first step is the acquisition of the SMCs by self-generation of sensory stimulation through interaction with the environment. The acquired SMCs need then to be utilized for control, which implies the incorporation of some normative framework for evaluation of actions and resulting sensory states. This corresponds to the learning of intention-related SMCs (or the SM strategy in Buhrmann's terms). The approach is different from the one presented in the previous chapter (Chapter 3): instead of inspecting the SMCs explicitly and how they are affected by several components of the brain-body-environment system, here the rules of the SMCs are learned directly and utilized for control. The SMCs can therefore be represented in abstract ways, specific to the machine learning techniques at hand. We start by employing the Markov model for SMCs introduced in Section 2.2.1 for control of the Puppy robot and later extend it by embedding it into a reinforcement learning framework and replacing its learning architecture by novel machine learning methods from reservoir computing. The work presented in this chapter is summarized in the publications [Hoffmann et al., 2012], appendix D and [Schmidt et al., 2014], appendix E.

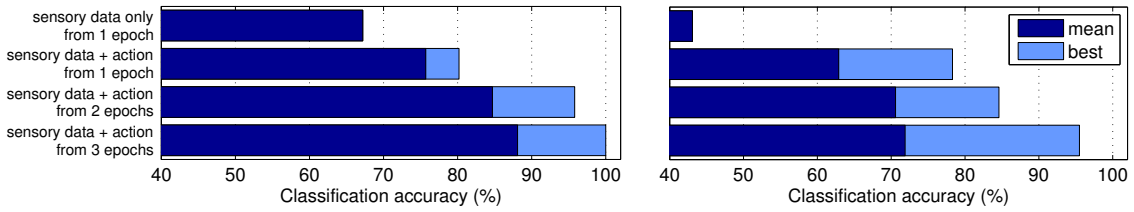
4.2 Markov model for learning SMCs

The Markov model for SMCs introduced in Section 2.2.1, published in its first version in [Maye and Engel, 2011], uses conditional probability distributions of action-outcome pairs (of variable history length) to store experienced SMCs. In [Hoffmann et al., 2012], appendix D, this learning method was applied to the Puppy robot to demonstrate its performance in a more complex embodiment. To this end, experiments were performed, in which the robot was equipped with a repertoire of five different gaits—sets of motor control parameters—(in accompanying simulations even nine) that it could chose from every two seconds. It ran in an arena with four ground-materials with different friction and surface structure and needed to select the right gaits (actions) that maximized its internal value system. The internal value was defined based on a stability measure, speed and tumbling. In the context of SMCs, the different terrains conceptually corre-

sponded to objects that the agent interacted with and which it needed to distinguish, in order to select appropriate actions for achieving its task.

One primary aim of this study was to investigate how classification performance is affected by incorporation of action information. The claim was that, when taking action context (efference copy as additional input to the classifier) explicitly into account, the performance of object classification increases compared to classification based on sensory information only. This was tested by comparing the performance, first based on the sensory part of the collected SMCs, and second on the whole SMCs as action-observation pairs. Furthermore, the effect of a longer sensorimotor context for classification was investigated with the aim to show that object-related SMCs live on expanded time scales (as suggested in [Maye and Engel, 2012a]). Following this claim, the expected outcome was an increased classification performance with growing history length of the SMCs (the conditional probability distributions forming the SM habitat). Finally, the study aimed at showing that the robot can successfully deploy its sensorimotor knowledge about the different terrains to choose appropriate gaits that maximize its locomotion stability, which corresponds to implicitly learning the intention-related SMCs (and thus a SM strategy).

The results of these experiments show that using this model the robot could indeed maximize its internal value by choosing appropriate actions according to the terrain it was currently walking on. This implies that it performed some implicit ground discrimination based on the SMCs it experienced before. Results further show that the robot was able to detect the current terrain it was walking on more effectively when taking not only the sensory information, but also the motor information into account, confirming the advantage of adopting the enactive approach on perception (Figure 4.1). Finally, terrain classification got even better with longer history length of the action-observation context, confirming the hypothesis that object-related SMCs live on longer time scales than modality-related SMCs.



(a) Real Robot (b) Simulated Robot

Figure 4.1: **Terrain classification in Puppy.** Comparison of classification accuracies when the action context and sensorimotor history are taken into account to different degrees. Figure taken from [Hoffmann et al., 2012].

Although this SMC model was successfully applied to the more complex and compliant Puppy robot, the sensory and motor spaces had to be drastically down-sampled and binned, in order to get moderately populated probability distributions (especially in the cases of longer history length). The representation as discretized probability distributions bore the problem that interpolations were hard to achieve (not to mention extrapolations). The employed action-selection policy was therefore to choose a random action whenever an unseen sensorimotor state was encountered, no matter whether the robot experienced similar, slightly different states before. This basically renders generalization impossible in this architecture. As a result, without the severe limitation in dimensionality and quantization, the sensorimotor space would have been so large

and so sparsely populated, that a matching of previously experienced SMCs could have occurred only very rarely in reasonable time, such that the robot would not have been able to exploit its accumulated knowledge properly, but would have nearly always selected random actions.

This drawback was in part compensated by the embodiment and situatedness of the agent which constrained the effective sensorimotor trajectories that were encountered—the SM habitat—a lot (in the case of the Puppy robot the space contained only 4% of all theoretically possible sensorimotor state combinations) and using sparse implementations of the distributions, the memory usage could be decreased to a moderate amount. While there is some effort to overcome the above mentioned shortcomings by efficient implementations of the probabilistic model using tree structures that improve the computational and memory complexity [Maye and Engel, 2012b], we further extended this model by replacing its core learning and storage architecture (conditional probability distributions) with more sophisticated machine learning methods. The next section will introduce our work in that direction.

4.3 Learning intention-related SMCs with reinforcement

The model described in the previous section is a straight-forward implementation of the SMC concept as articulated in its original publication by O'Regan and Noë [2001] and has shown to be a promising model for controlling artificial agents. This section presents how the shortcomings mentioned above, especially the generalization limitation for higher-dimensional continuous systems, can be overcome by employing suitable machine learning architectures.

4.3.1 The actor-critic design in reinforcement learning

The probabilistic Markov model connects the gathered action-outcome knowledge to an internal value that is used to evaluate the outcomes and select those actions that have highest probability to result in a high internal value. This is one major concept in reinforcement learning and in order to improve the existing model it seems valuable to reformulate it in terms of this well studied framework, which provides methods for many different types of problems.

Similar to the internal values defined in [Hoffmann et al., 2012], appendix D, in reinforcement learning, each transition from one state to the next mediated by a certain action is rewarded or punished. The agent's goal is to learn a control policy that leads to maximal accumulation of rewards and minimal receiving of punishments. A key part of this strategy is the prediction of future rewards r_{t+1+k}, given a certain action a_t and a certain state s_t, the so-called action-value function:

$$Q(s_t, a_t) = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \cdots = \sum_{k=0}^{\infty} \gamma^k r_{t+1+k} \tag{4.1}$$

where γ is a discount factor, weighting the future rewards. The classical approach to learn this prediction is to visit all possible actions from each possible state and then build a look-up table. However, this is not feasible for most real-world situated agents, because of the combinatorial explosion of the number of possibilities, and even more so as a history of several action-outcome pairs is considered. But there exist reinforcement learning methods that approximate Q from already experienced agent-environment interaction. This can be done even on-line using the so-called temporal difference (TD) error [Sutton and Barto, 1998] that is built from the state, action and reward observations of two successive time steps:

$$E_{TD} = Q(s_t, a_t) - [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1})] \tag{4.2}$$

This error term can be used as a teaching signal for supervised learning of the action-value function in an iterative (on-line) manner. Since it uses the prediction of one time step to correct the prediction of the previous one, it is a bootstrapping method that will approach the optimal solution only after having sampled parts of the state-action space several times. The algorithm employing the TD error is called *SARSA* (named after the employed variables $s_t, a_t, r_{t+1}, s_{t+1}, a_{t+1}$) and has a deviate for continuous-valued state-action spaces, the so-called actor-critic design (ACD) [Sutton and Barto, 1998]. In short, in ACD, the selection of actions is done by an actor module, which is separated from the estimation of Q performed by the critic module. The critic needs a function approximator that is capable of learning Q using the TD-error as teaching signal. The actor needs to be capable of utilizing the critic's estimate for selecting optimal actions. In the present work, the actor was simply gradient ascending the estimated Q function to find the action that maximizes it (see [Schmidt et al., 2014], appendix E for further details). We will discuss more advanced possible actor networks in Chapter 6.

Compared to the probabilistic SMC model used in the previous section, here the learning problem is shifted from a forward model-like prediction of future sensory states (corresponding to modality-related SMCs) and then connecting them to an internal value or reward, towards directly predicting the future rewards that will be received when following the current action in the current state. This corresponds to direct learning the intention-related SMCs.

4.3.2 Reservoir computing for actor-critic

As mentioned before, the actor-critic design needs a function approximator to learn to predict the action-value function Q from the current sensorimotor state. Artificial recurrent neural networks are promising candidates for replacing the probability distributions for capturing SMCs, due to their capability to approximate theoretically any dynamical system in continuous spaces. Particularly, echo state networks (ESN) [Jaeger, 2001] have gained wide attention due to their new architecture, efficient training and successful application to a variety of problems. The key concept is that the recurrent part of the network, the *reservoir*, which receives the input, is initialized with a large number of randomly and sparsely connected neurons, whose recurrent connection weights remain fixed throughout learning. The only weights that are subject to training are those of the readout connections that connect the reservoir neurons to an output. The recurrent reservoir can thus be seen as a high-dimensional, non-linear and temporal expansion of the input space. In the high-dimensional feature space the desired function can be learned in a supervised manner by combining and weighting the features (e.g., linearly) to produce the desired outcome (e.g., using any standard regression method).

In contrast to probability distributions used in the Markov SMC model, the ESN-critic does not store all observed state transitions. Instead, the reservoir activation represents a trace of the recently experienced sensorimotor stream, which is fading over time and whose relation to future rewards is remembered in form of the readout connection weights. This breaks the scalability limitation by reducing the memory complexity to input weight matrices, which grow linearly with increasing input dimensions. Furthermore, ESNs naturally interpolate between learned inputs/outputs and can even extrapolate to a certain extend, thus having overall good generalization capabilities. Last but not least, being artificial neural networks, they constitute a biologically plausible implementation of the learning process.

Recently, the successful usage of ESNs in an actor-critic design has been demonstrated in an obstacle avoidance task of a simple mobile robot (ePuck) [Koprinkova-Hristova et al., 2010; Oubbati et al., 2012]. Here we investigate how the method copes with potentially high-dimensional, compliant robots with more complex dynamics.

4.3.3 A case study in the Puppy robot

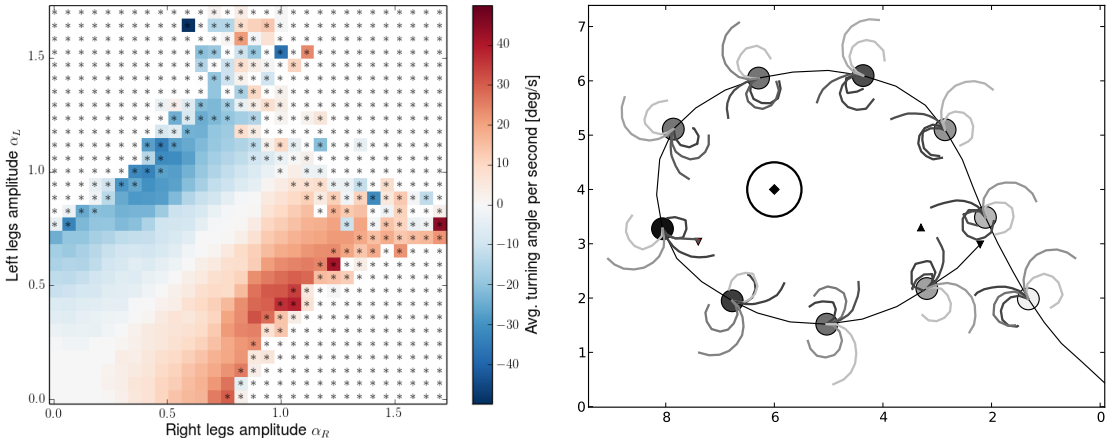
In a case study, we implemented the actor-critic design using echo state networks (ESN-ACD) for the simulated version of the Puppy robot, having the advantage to be able to investigate the system's dynamics and learning properties in full detail. The scenario was a navigation task, where the robot had to walk towards a target location. The robot was controlled by a sine wave pattern generator, implementing a bounding gait. Navigation could be achieved by adjusting two parameters of the sine waves, controlling the speeds of the two left and the two right legs, respectively (resulting in a two-dimensional continuous action space). Intuitively, larger speeds on the right legs than on the left legs led to leftwards turning and vice versa, but the actual mapping was more complex and non-linear, as can be seen in Figure 4.2a. Especially, certain combinations of left and right speeds could make the robot fall and the behavior was critically depending on previous actions.

First experiments of a simplified simulation setup showed successful training of the critic [Baumgartner, 2013]. For simplicity, instead of using any raw sensor signals, here we used the distances to several artificial landmarks in the arena as sensory states. The task was mediated by the reward function, which was defined as the distance to the target location. The action space was constrained to regions where the robot did not tumble, limiting its speed and agility. Figure 4.2b shows the critic's performance after training. In this analysis, the robot was enforced to use a predefined action sequence. At each control step (gray disks), the action-value function Q was estimated for several possible actions. The outgoing gray lines indicate the trajectories the robot would have walked if it had been executing certain actions (other than the enforced actions) and color intensities indicate the estimated Q-value for the respective actions. It can be seen that the lines pointing towards the target location (black circle in the center) have higher Q (darker colors) than those pointing away from the target. This confirms that the robot had learned successfully to estimate the future rewards and that, would it not have been enforced to the predefined actions but to choose its actions based on those estimates, it would indeed have walked towards the target.

In [Schmidt et al., 2014], appendix E, we extended these experiments to incorporate raw sensor readings of the robot as inputs to the ESN-critic, without any preprocessing, such as downsampling, down-quantization or manual feature extraction. Here, the robot's compass sensor was used, a 3D vector pointing to the virtual north (resulting in a 3-dimensional continuous sensory space).[1] Furthermore, the reward function was designed to not only reward heading towards the target, but also to punish tumbling of the robot. The constraints on the action space were removed and the robot had to learn to avoid actions that resulted in tumbling on its own. The action selection followed an ϵ-greedy policy, where ϵ was the probability to select a random action and $(1 - \epsilon)$ was the probability to select the actor's recommended action. The ϵ had a high value in the beginning and was slowly reduced during the experiments, letting the actor take over control more and more over time, as the critic improved its estimation of intention-related SMCs.

The results, summarized in Figure 4.3, show that the ESN-ACD model was indeed able to learn the intention-related SMCs in reasonable time and utilize them for selecting reward-maximizing actions (avoid tumbling, walk towards target). This was despite the high complexity of the robots interaction with its environment and the continuous-valued state and action spaces. Figure 4.3a shows that, after training, the critic predicted highest action-values when the robot headed towards south east, where the target was located (thus having the compass-north in its rear-left direction). Also, those compass values occurred much more often than other directions, indicating that the actor managed to walk the robot into that direction most of the time. This is confirmed

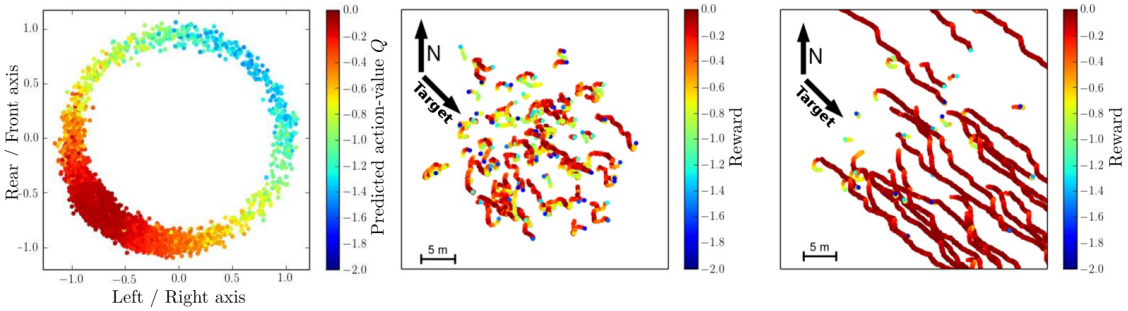
[1] For navigation on a 2D surface a 2D vector might seem sufficient. However, since the robot was pitching and rolling a lot during locomotion, the third dimension of the compass sensor provided information about the robot's current orientation in 3D space.



(a) Action-behavior mapping

(b) Critic evaluation of different actions

Figure 4.2: **Behavior of Puppy resulting from different actions. (a)** Action-behavior mapping for actions performed after running straight ahead for some time ($\alpha_L = \alpha_R = 0.5$). The color indicates the turning angle of the robot when performing the respective action. Actions that led to tumbling of the robot are marked with an asterisk character. Figure taken from [Schmidt et al., 2014]. **(b)** Critic evaluation on a set of predefined actions in a target-navigation task. The target location is indicated by the black diamond marker surrounded by a circle. The gray discs represent the robot's position with color intensity indicating the current predicted action-value Q. The outgoing lines show the effect of certain actions on the trajectory with the color intensity indicating the value of Q when the robot would execute the respective action. The darker the color, the higher the prediction of Q. Figure taken from [Baumgartner, 2013].



(a) Compass sensor vs. predicted action-value

(b) Trajectories before learning

(c) Trajectories after learning

Figure 4.3: **Performance of actor-critic design in the navigating Puppy robot. (a)** The compass sensor readings after training with the color indicating the predicted action-value function Q. **(b,c)** The trajectories the robot runs before (b) and after (c) training with the color indicating the actual reward received at each time step. In this case the target was located in the south-east direction. Figure adapted from [Schmidt et al., 2014].

by the trajectory plots (Figure 4.3b,c) that show short random walks before training (tumbling occurring often) and long straight walks towards the target after training converged, yielding high rewards over long periods.

4.4 Summary

In this chapter, we investigated SMCs in the context of task-oriented learning. This comprised not only the capturing of the SM habitat including modality-related and object-related SMCs for learning about the agent's morphology, the structure of its sensors and motors, or about its environmental structure such as objects. It moreover involved a normative framework by learning intention-related SMCs (the SM strategy) that yielded information about how certain actions affected the internal value or accumulated rewards when the agent was in a certain state.

Starting from the probabilistic Markov SMC model, we showed that SMCs could be utilized to detect objects (corresponding to the terrains the robot was walking on) and to select actions that increase the internal value. Importantly, this was achieved from purely self-generated sensorimotor experience and no symbolic labels had to be defined, stored or recalled. The agent rather identified the objects by actively interacting with them which induced specifically structured sensorimotor streams that can be seen as object-specific "fingerprints". Furthermore, it was shown how the detection accuracies improved, the longer the interaction with the specific object lasted. This corresponds to the concept by O'Regan and Noë [2001] of perceiving by persistently exercising the SMCs through interaction.

Despite the achieved results, we identified some limitations of the probabilistic Markov SMC model regarding scalability and generalization capabilities. We demonstrated how those limitations could be overcome by embedding the model into existing reinforcement learning frameworks and employing new machine learning techniques. ESN provide natural interpolation (and even to some degree extrapolation) and have a reasonably growing computational and memory complexity with increasing number of input dimensions. Using the ESN-ACD method, the robot could achieve its tasks using raw, unprocessed sensor readings, which made manual feature design dispensable. This way, the knowledge that was externally put into the system was reduced to a minimum amount. The only information that had to be provided by the experimenter was the reward function (and some parameters of the learning machine). However, the reward function was crucial for the behavior and the task performance. It had to be designed carefully and had to match the information from sensors available to the agent. The critic needed to be able to derive future rewards from the sensor values and actions. If, however, the sensors would not provide enough information about the task performance, the critic could do no better than averaging rewards experienced at certain actions, which might not be a very powerful prediction.

Chapter 5

Models of Human SMCs for Artificial Systems

This chapter studies computational models of human SMC learning using data obtained from psychophysical experiments to derive corresponding learning strategies for artificial agents. First, by discussing [Kootstra et al., 2012], appendix F, it is shown how a combination of internal forward and inverse models can not only reveal the processes behind human adaptation, but furthermore provides a method that allows artificial systems to learn their SMCs and to self-adapt to altered SMCs. Second, as presented in [Wilming et al., 2013], appendix G, the saliency-based action selection strategies of human eye-motion are modeled mathematically, which provides a basis for autonomous systems to efficiently sample relevant parts of a scene.

5.1 An adaptation model for artificial agents

Humans can adapt quickly to disruptions of sensorimotor coordination, as was impressively demonstrated in prism-adaptation studies, where subjects were exposed to prism glasses that displaced their visual field horizontally by several degrees [Kornheiser, 1976; Redding and Wallace, 2006]. In those situations, the SMCs are altered drastically, but humans still manage to adapt their movements and restore eye-hand coordination after performing a few movements under concurrent visual and proprioceptive feedback. Besides the insights for the study of human learning, such mechanisms would be highly demanded for artificial systems such as robots. Those systems often need to be calibrated before use and moreover, are subject to mechanical wear and use or failure. These effects change the SMCs of an agent and render formerly learned control schemes obsolete.

5.1.1 Learning with forward and inverse kinematic models

The method presented here (see [Kootstra et al., 2012], appendix F for details), uses a combination of forward and inverse kinematic models, to learn movements of a simulated robotic arm. We used this setup to reproduce the behavioral results of a psychophysical study investigating the adaptation to prism glasses in humans [Redding and Wallace, 2006]. Through this transformation of the visual field, the lawful relations between sensory and motor signals—the SMCs—were changed and the agent had to re-adapt in order to maintain performance.

In model learning, what is usually required is a functional mapping. For instance, when the task is to move the arm of a robot to a desired position (e.g., in order to perform some manipu-

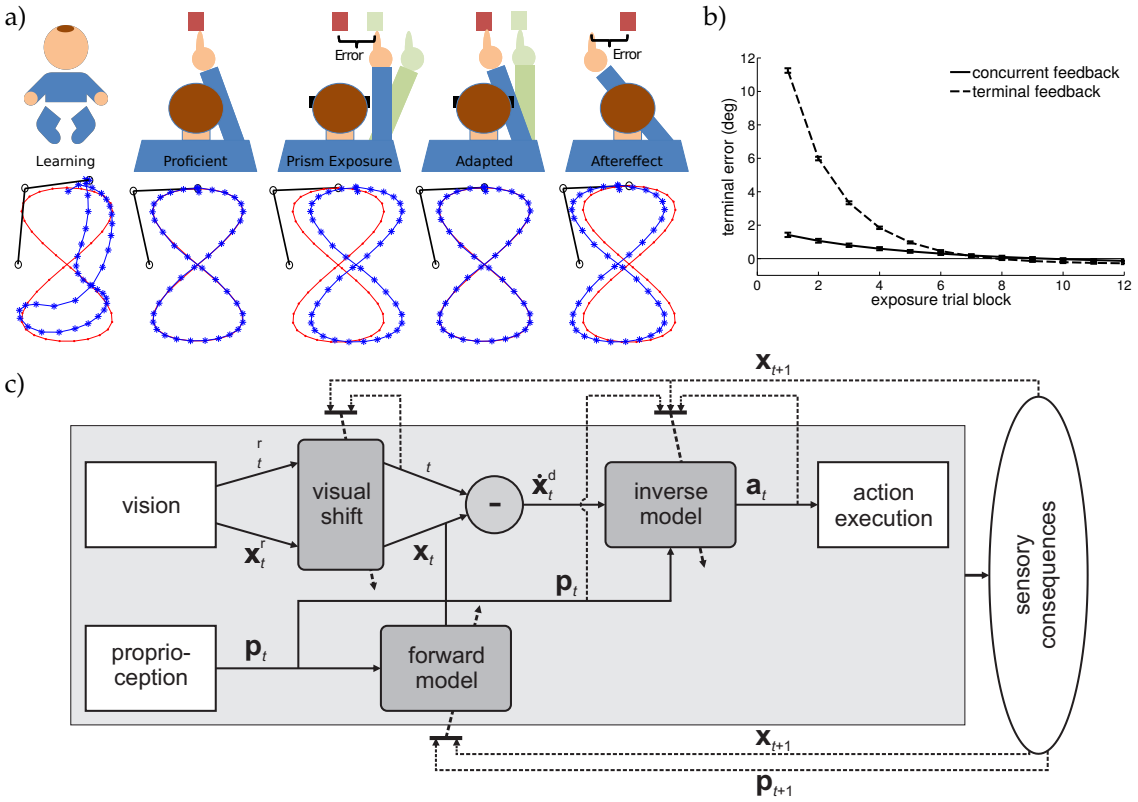


Figure 5.1: **Prism adaptation experiments in a simulated robotic arm. (a)** Schematic of the learning and adaptation process in humans and in our model system during prism-adaptation experiments (from left to right). The model learned from scratch to follow a desired trajectory with its end-effector. When exposed to prism glasses, the visual field was shifted, which resulted in reaching errors. After adaptation, performance was restored, but an aftereffect could be observed when prisms were removed again. **(b)** Pointing error after applying the prisms (camera shift) in two different feedback conditions: With concurrent feedback (solid line) the arm was visible throughout the trial, whereas with terminal feedback (dashed line) the arm became visible at the end of the trial only and the model had to rely on the proprioceptive forward model. **(c)** Theoretical model, including three sub-modules: the visual shift, the forward model, and the inverse model. Based on visual and proprioceptive observations, the model estimated the action necessary to move the end effector to the target pose. All three sub-modules learned on-line using the sensory consequences of the executed action as observed by the model itself. Information flow for control is indicated by solid arrows, and the flow for learning by dashed arrows. Figures (b) and (c) taken from [Kootstra et al., 2012].

lation), a mapping is needed from the desired change in hand position within the robot's camera image to the change in angles that needs to be applied to the shoulder and elbow joints in order to reach the desired position. Such a *differential inverse kinematic model* can be learned from own experience, by first randomly moving the arm, and then combining the sampled joint angle changes and observations in the camera image as training examples in a supervised learning method [see e.g., Nguyen-Tuong et al., 2009; Nguyen-Tuong and Peters, 2011; Baranes and Oudeyer, 2012]. Machine learning provides many sophisticated regression methods for such a function approximation [see Sigaud et al., 2011, for a comparison of methods]. Here we used an on-line version of Gaussian process (GP) regression [Rasmussen and Williams, 2006] to learn the mapping. Because visual feedback was not available in all experimental conditions, we incorporated a second GP regression for learning the forward kinematic model for the proprioceptive system in a similar way. In the forward model, the mapping returns the assumed position of the arm's end-effector in the camera image, given the current joint angles. The setup and the theoretical model are shown in Figure 5.1. Going from left to right of Figure 5.1a, the experiments consisted of first learning to follow a target trajectory (requiring some random exploration in the beginning while training the GPs). After the agent had successfully learned the task, it was exposed to the prisms (simulated by a visual shift of the camera image) and the adaptation behavior was measured. When the system had adapted, the prisms were removed and the aftereffect was investigated.

Figure 5.1b shows how the system adapted after simulated exposure to the prisms. With terminal feedback (the arm position was only visible at the end of a trial), the pointing error started close to eleven degrees, which corresponded to the shift in the camera image, and decreased to perfect pointing performance after 8 trials. With concurrent feedback, the arm was visible during the whole trial, which allowed for corrections during pointing and resulted in lower pointing errors from the beginning.

5.2 A model for efficient visual exploration

The study discussed in this section ([Wilming et al., 2013], appendix G) aims at understanding the mechanisms behind human control of eye movements and how those could be translated into an attention model for artificial agents. Eye movements serve the purpose to concentrate processing resources on relevant stimuli and gather new information. With respect to sensorimotor contingencies, this seems complementary: SMCs are defined by the lawful relation of action and sensory changes, which can be nailed down to prediction of future sensory information. The question arises how the novelty components of the bottom-up signals compare to and match with the predictive components that arise from learned SMCs. The main conclusion points to an intrinsic trade-off between exploration and exploitation of information about the visual scenery. With respect to artificial intelligence this is an interesting feature that would allow robots to efficiently sample the visual field and at the same time focus their attention on relevant parts of the scene. Existing robot models of visual attention usually rely on intensity- or motion-based saliency, sometimes coupled with mechanisms inhibiting previously visited locations [Ruesch et al., 2008; Begum and Karray, 2011].

5.2.1 The study of *inhibition of return*

Our study investigates a heavily debated phenomenon called *inhibition of return* (IOR) which refers to the observation that, when humans look at the same location twice, i.e., return to a location after having visited another *out-location*, they seem to spend more time than usual at the out-location before they come back to the return-location [Posner and Cohen, 1984]. The increased

latency to re-fixate a given location hints on some kind of inhibition mechanism. Another observation indicates a decreased probability to re-fixate a location at all [Klein and MacInnes, 1999]. These effects were believed to serve a foraging strategy, facilitating exploration of the visual scene. Therefore, IOR reflects a part of the decision process between sampling new information or extended processing of known visual areas.

In [Wilming et al., 2013], appendix G, we analyzed a large dataset from eye-tracking experiments, where ocular movement was recorded during a free-viewing task. The dataset from 5 different studies comprised more than 550,000 fixations from 235 subjects looking at natural images, urban images, fractals and pink noise. An example of return saccades with one or two intermediate out-locations is shown in Figure 5.2a. Each fixation was measured by the duration the subject fixated a patch in the image in combination with the length of a saccade and its relative angle w.r.t. the previous saccade. (Figure 5.2b). We analyzed the frequency of return saccades and compared them to estimates of the number of return saccades expected from statistical properties. We furthermore investigated temporal properties of return saccades (fixation durations at preceding fixation locations) and compared them systematically to those of forward saccades. Finally, we analyzed the fixation data w.r.t. bottom-up saliency, defined by local image properties.

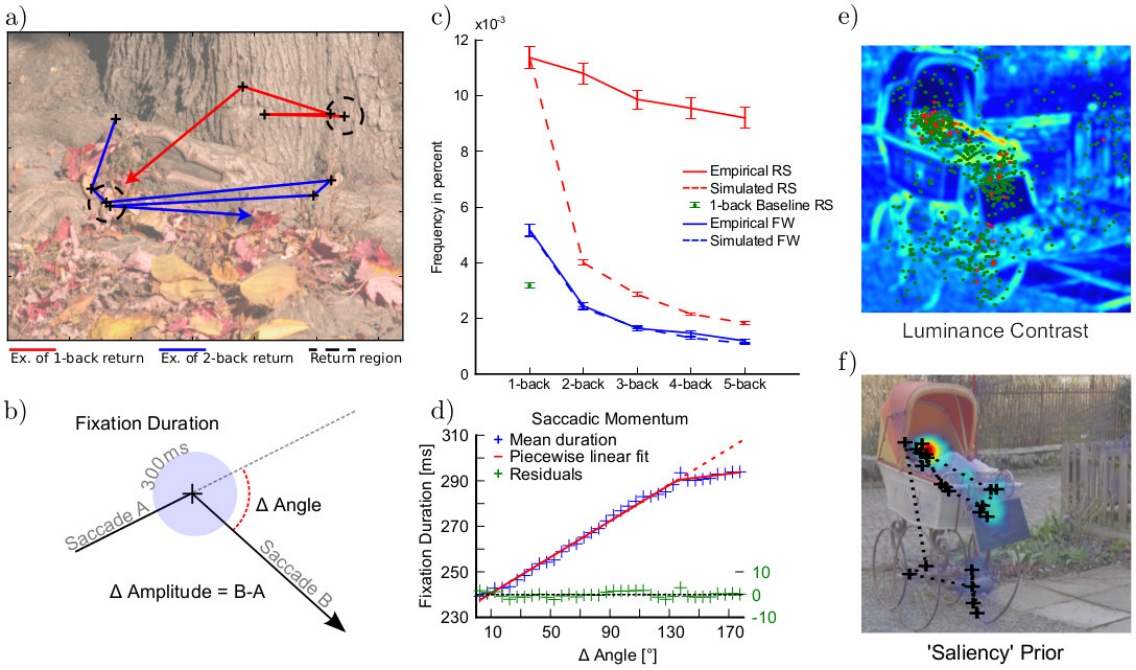


Figure 5.2: **The study of return saccades. (a)** Examples of 1-back (red) and 2-back (blue) return saccades. **(b)** Fixations were quantified by their duration at a certain location, the angle of movement relative to the previous direction and by saccade length. **(c)** Spatial frequency of return saccades (red) vs. forward saccades (blue). Empirical data (solid lines) compared to statistical expectation (dashed lines). The green bar corresponds to the chance baseline. **(d)** Fixation durations as a function of saccade angle relative to the previous saccade (blue). The red line shows a piecewise linear fit and the residuals from this fit are shown in green. **(e)** Example of a visual feature (luminance contrast) of an urban image. Regular fixations are shown as green dots, return fixations in red. **(f)** Fixation probability of the same image based on saliency. Black crosses and dotted lines show an example trajectory. Figure adapted from [Wilming et al., 2013].

Spatial inhibition of return

Figure 5.2c shows the empirically measured frequency of return and forward saccades in comparison to simulated saccades based on the statistical distribution of relative saccade angles and lengths. In contrast to the prevalent belief in the research community that the regions that subjects returned to have lower probabilities to be re-fixated, we found the opposite, a higher fixation probability. This holds equally for return fixations with only one intermediate fixation before returning, as well as for those with multiple out-locations. The data suggest a facilitation of returning to previously fixated locations, rather than an inhibition of return.

Temporal inhibition of return

Measuring the latencies of fixations preceding a return saccade (Figure 5.2d) revealed that the known effect for temporal inhibition of return can be fully explained in terms of another known effect, the so-called "saccadic momentum" [Smith and Henderson, 2009]. This effect denotes a linear increase in latency with higher turning angle, which is not specific to the re-fixation of certain locations, but holds for all saccades. Once this momentum is taken into account, there is no increased latency before returning, but actually an increased latency at the location that is re-fixated. This indicates extended processing demands at the location that is visited multiple times.

Bottom-up saliency

The increased fixation duration at return locations suggests that those locations are special. To clarify the role of locations, we compared bottom-up saliency at return locations and regular fixation locations. We used a model of bottom-up saliency derived in [Wilming et al., 2011], which is based on a weighted sum of low- and mid-level visual features (Figure 5.2e). Results reveal that return locations are better predicted by those features than regions that are fixated only once, suggesting that return saccades are directed to more salient locations than normal saccades.

5.2.2 A bottom-up sampling strategy

The increased number of return saccades and the prolonged fixation durations at return locations are difficult to reconcile with a foraging strategy that maximizes the visited area by distributing fixation locations uniformly in the visual field. Instead of such a maximization of entropy, we hypothesized that return fixations optimize the coverage of an internal relevance map of the visual scene. We calculated such an image-specific map from the empirically measured saliency of all image locations, which was defined by how often subjects looked at locations in the particular image. Our analysis shows that allowing for return fixations increases the probability to cover the relevance map compared to forcing all saccades to unexplored regions.

The study presented here thus reveals an important aspect of our strategy to select fixation points in our visual field: that we trade off exploring our environment against making sure that we have fully comprehended the relevant parts of it. Although the work is limited to the study of human eye movements, it provides a promising basis for attention models for robotic systems. Applying the saliency model to a camera image and combining it with the saccadic momentum model derived here, one could guide camera movements of a robot in order to efficiently sample its environment while keeping track of regions of special interest. In fact, similar models of visual attention for robotic cognition exist [Begum and Karray, 2011], many of them inspired by human models [see e.g., Itti and Koch, 2001] and some even include inhibitory mechanisms simi-

lar to IOR [Ruesch et al., 2008]. Our results can contribute to the refinement of such models w.r.t. optimal sampling.

5.3 Summary

The two studies presented in this chapter show how computational modeling of human behavior can inform control of artificial agents.

The results of the first study reproduced the adaptation effects observed in the psychophysical experiments and added some novel insights about the learning mechanisms in the human. More details can be found in the full paper, but importantly for the discussion about artificial agents, the GP learning scheme provided first, a system that can learn desired reaching movements from scratch by random exploration, i.e., without the need to specify a model of the plant or providing other prior knowledge to the system. Second, an adaptive system is provided that can cope with the changing conditions that lead to altered SMCs (transformation on the camera system in this case). The work demonstrates that adaptation can be seen as an active process that crucially depends on the ability to relate motor actions to their sensory consequences, which bears a resemblance to the concept of SMCs.

The second study investigated the processes behind selection of saccade movements. It was shown that the classical assumption of an inhibition process that serves the facilitation of foraging cannot fully account for the observations from the psychophysical experiments. Instead, the saccadic momentum effect, in combination with bottom-up saliency can explain the observed occurrence of return saccades in terms of a sampling strategy that optimally balances the exploration of the visual scenery and the processing of regions of interest.

Chapter 6

Discussion & Conclusion

6.1 Discussion of results

In this thesis, we investigated the sensorimotor contingencies theory from the perspective of artificial intelligence. We analyzed SMCs in artificial agents and how they can facilitate the learning process and the formation of cognitive behavior. In this section, we wrap up the presented results.

6.1.1 Information-theoretic analysis of SMCs

In Chapter 3, we analyzed the sensorimotor signals in a quadruped robot running on multiple terrains using tools from information theory. We showed that information transfer quantified by transfer entropy can reveal the non-linear relations among the sensory and motor signals of the robot. Those relations the SMCs of the robot were shown to be affected by changes to at least two components of the brain-body-environment system. First, by comparing the information structure during different control policies we found that the information transfer shows patterns specific to the gait that was used. Overall information transfer was enhanced when using coordinated motor signals that led to stable periodic movement patterns, whereas using random motor commands, overall information transfer was low, but very specific to the robot's mechanical topology. Second, when comparing information structures during running in different environments, namely different terrains, we also found patterns specific to each terrain type, which we attributed to identifying the object-related SMC.

We demonstrated that information structure can reveal the modality-related SMCs of the robot by comparing information transfer patterns of each sensory signal with those of other sensory signals, which results in a primitive form of a body schema. It is further possible to find the SM environment of the robot through motor babbling, by averaging the information transfers over all possible control commands, a way of marginalizing out the "brain"-part of the brain-body-environment system.

In our analysis, the information-theoretic approach has led to perceptual capabilities of the agent, but did not find its way into control strategies so far. However, the fact that information transfer was much higher when coordinated motor commands were used, suggests that information might be a driving force for learning or evolution of internal dynamics and morphology of an agent. Similar ideas have been pursued in [Sporns and Lungarella, 2006; Prokopenko et al., 2006], where coordinated behavior was achieved by evolving control networks with information-based fitness functions. Several information-theoretic approaches were proposed that use information directly for task-independent control. *Empowerment* quantifies the influence an agent has over

perceived parts of the world in information-theoretic terms [Klyubin et al., 2005b,a], for which an empowerment-greedy action selection strategy has been derived [Jung et al., 2011; Salge et al., 2014]. *Predictive information* measures the mutual information between past and future sensory states and has been translated into an gradient-based rule to update the control network [Ay et al., 2008; Martius et al., 2013]. Recent studies even combine the task-independent information-based control with task-oriented reinforcement learning [Zahedi et al., 2013].

An investigation of morphological effects on the information structure remains to be done. Related work in this direction suggest that information structure is strongly affected by the morphology and that information-increasing changes to the morphology can enhance information processing [Lungarella and Sporns, 2006]. Following the ideas of the *information parsimony principle* [Polani et al., 2007; Polani, 2009], which suggests that information can be seen as a natural resource that evolution of biological agents might optimize, it would be interesting to investigate, whether information-driven evolution can create morphologies that are optimal in some respect and advantageous for the agent's behavior.

6.1.2 Task-oriented learning of SMCs

The work presented in Chapter 4 goes beyond the unsupervised learning and analysis of SMCs. Here we aimed at learning control of the robot in order to achieve certain tasks. We did this by providing feedback to the robot about its task performance in form of an internal value or reward system.

We again followed the bottom-up approach by capturing sensorimotor relations from random interactions, which were used as knowledge base for later task-oriented control. However, this time we additionally provided higher-level teaching signals that allow the formation of intention-related SMCs. The intention-related SMCs then enabled the robot to appropriately select actions that achieve the task.

Starting with the Markov model for SMCs, we showed how probability distributions can be used to learn the action-outcome relations of the quadruped Puppy robot and allowed it to discriminate several terrain types it was running on (similar to the information structure discussed above). A major contribution was to demonstrate how the incorporation of action information can enhance the classification process. Terrain classification accuracy was larger when using sensory information in combination with action information than when using sensory information only. Results showed further that sensorimotor relations observed over extended time spans can improve the classification even more. This suggests that object-related SMCs manifest themselves in a longer sensorimotor context than modality-related SMCs do. Here our work complements previous investigation of this subject [Maye and Engel, 2012a]. The modality-related SMCs are instantaneous relations with specific characteristics, whereas in order to recognize objects, a certain time of interaction is required. Imagine the difference of blindly grasping a spoon or a fork, which might feel very similar in the first moment, but when "playing around" with them in the hand for some seconds, we can easily feel the difference. A similar effect was observed in [Hogman et al., 2013], where objects were pushed in several different ways by a robotic arm and the observed action-outcome relations were used to classify the objects. Their results confirm that classification accuracy increases with the number of pushes performed.

After identifying some limitations of the Markov model in terms of generalization and practical usage, we proposed to further develop this approach while rephrasing the learning problem in the language of reinforcement learning and employing state-of-the-art machine learning techniques from reservoir computing. Therefore, we implemented the actor-critic design for the Puppy robot, which uses the TD-error to directly learn the intention-related SMCs in form of a mapping of sensorimotor signals to expected future rewards. We used echo-state networks as

function approximator, which have certain advantages for our purposes. First, ESNs are universal approximators of dynamical systems with good generalization abilities that can deal with noisy or previously unseen data. Second, the recurrent reservoir allows capturing SMCs on several time scales, ranging from instantaneous relations such as modality-related SMCs to longer time scales where intention-related SMCs are supposed to live. Especially in compliant robots this memory capacity is of great importance for capturing the dynamics of the system. Third, for the gradient ascent method used in our approach to select reward-maximizing actions, ESNs provide an efficient way of calculating partial derivatives of outputs (future rewards) with respect to the inputs (actions). We demonstrated that this combination of methods could successfully learn the complex action-intention relations of the compliant Puppy robot and enabled it to maximize rewards by appropriate selection of actions. This way, the simulated robot was able to learn to autonomously navigate towards target locations, which were mediated by the reward function.

In this setup, the reward function is the only externally provided knowledge source of the system, all remaining knowledge about successful control is learned from own experience by the interaction with the environment. This means, however, that the design of the reward function is crucial for the entire behavior of the robot. It has to match the morphology of the robot with its specific possibilities of actuation and its specific sets of sensors, while at the same time encoding the robot's task.

Both presented SMC models require random exploration which bears the problem of how to efficiently sample the sensorimotor space. We will discuss possible strategies in Section 6.3.

6.1.3 SMCs in model learning

In Chapter 5, Section 5.1, we looked at SMCs from a different perspective. Here we investigated how an artificial agent can cope with morphological changes that alter its SMCs. We provided a model that is capable of adapting to the changed SMCs in order to restore its functionality. The model comprised a forward and an inverse kinematic model learned by Gaussian process regression. By moving its actuators and observing the resulting sensory changes, the agent used the captured modality-related SMCs to train the two models iteratively, similar than in the reinforcement learning studies. In contrast to those, however, no intentions were introduced, but a mapping was derived instead that allowed direct, low-level control of the agent in order to achieve desired sensory states. In Buhrmann's terms this corresponds to the domain of SM coordinations.

The model was used in a simulated robotic arm that had to point to desired target locations. In our study, we aimed at reproducing pointing behavior of humans when wearing prism glasses that displace their visual fields. In such situations, humans can quickly adapt to the new SMCs imposed by the prisms after practicing for a few trials. While our experiments with the robotic arm successfully reproduced those results, they support the pragmatic stance of cognition by revealing that this adaptation mechanism is an active process: it requires active engagement in the task in order to learn the changed SMCs.

With respect to AI and robotics, such a capability to self-adapt is highly relevant. Robots are often subject to mechanical wear or are used in changing environments and need to be calibrated before being fully operational. A system that can learn its changed SMCs by moving around a bit is thus of great value. Furthermore, in such a system, the usage of tools can be learned autonomously by learning the new arising SMCs (cf. body schema extension in [Hoffmann et al., 2010]).

By implementing forward and inverse models, the study further provides a biologically plausible approach to learning and adaptation mechanisms, as the abundant work in neuroscience and psychology about internal models in the human brain indicates [see e.g., Wolpert et al., 1998; Bays and Wolpert, 2007; Imamizu and Kawato, 2012].

There exists a difficulty in control of robotic arms that becomes especially apparent with larger number of joints (degrees of freedom). While the learned mapping in the small vicinity of the current arm configuration might be unique, there might be multiple possible joint trajectories to move the arm's end-effector to target positions that are further away. This redundancy issue has been recently tackled in an approach called *goal-babbling*, by weighting the training samples based on how close the movement brings the hand towards the target [Rolf et al., 2011; Hartmann et al., 2012]. Other approaches use alternate freezing and freeing of degrees of freedom in order to overcome this difficulty [Berthouze and Lungarella, 2004], or a from of intrinsic motivation to guide action selection [Baranes and Oudeyer, 2013] (see also Section 6.3).

Our last presented study of this thesis (Section 5.2) was concerned with the involvement of modality-related SMCs in visual exploration. By analyzing a large set of eye-tracking data where humans were engaged in a free-viewing task, the spatial and temporal effects of "inhibition of return" were investigated, which presumably inhibits returning to previously seen locations in the visual scene, in order to facilitate foraging. Our results revealed that no such effect exists, but rather a "saccadic momentum" effect, which slows down saccades gradually with deviation of their direction from the direction of previous saccades. Accounting for this effect, our results further showed that returning to previously seen locations is actually enhanced and that those locations we return to are more salient that other locations. We conclude this study by suggesting that the saccadic momentum effect implements an optimal sampling strategy of salient regions of the visual scenery, while trading off exploration of unseen locations. We substantiated this claim by simulations of eye trajectories using the statistics obtained from the human eye-tracking experiments. These simulations showed indeed higher likelihood of the empirically measured image coverage when allowing return-saccades than when removing them from the data set.

Such an optimal sampling strategy could be implemented in robot visual control systems to improve attentional mechanisms. Existing models of robot visual attention indeed use saliency-based focusing of eye (camera) gaze [Begum and Karray, 2011]. There are even models accounting for what is widely termed inhibition of return: a prevention mechanism to refocus previously attended regions in the visual field [e.g., Ruesch et al., 2008]. However, with respect to optimal sampling and trading off exploration and exploitation of the visual scenery, our findings might complement and improve existing models.

Our model comes with the limitation that it is bound to static images and does not include movements in the visual scene. The human visual system, however, incorporates such dynamic information in attention mechanisms [e.g., Itti and Koch, 2001]. This could be accounted for by considering dynamic features in the saliency evaluation of images and image sequences, additionally to the static image features that have been considered here.

6.2 Categories of SMCs

In Section 1.2.2 we introduces three notions of SMCs: modality-related, object-related, and intention-related. Buhrmann et al. [2013], on the other hand, defined four regimes of SMCs from a dynamical systems perspective: sensorimotor environment, habitat, coordination and strategy. Which classification makes more sense? How are classifications of SMCs useful at all?

In the original publication [O'Regan and Noë, 2001], the term "Sensorimotor Contingencies" embraces all relations between movements and accompanying changes in the sensory system. The authors point out that these relations are what our sensations are made of. Because the effects of movements on sensory signals coming from, let's say, the auditory system follow different rules—are differently structured—than the effects on visual sensory signals, we can differentiate and be aware of the two modalities. It thus seems valuable to use the notion of modality-related

SMCs in order to identify modalities in the sensorimotor stream. This distinction enabled us to build a rudimentary body schema in Section 3.3, which can be useful to an agent not knowing anything about its body and the nature of its sensors. Similarly, the sensorimotor relations differ when interacting with different environmental features, e.g., objects. For many applications of artificial agents the capability to distinguish objects is of great importance. Finally, specifying the rules of how actions affect the internal value system as intention-related SMCs seems useful for learning strategies like reinforcement learning. This last type can be seen as an extension of the SMC concept described in the original publication to higher cognitive levels. The action is generalized from mere movements to actions, which might include action sequences or sets of actions and the outcome is not sensory anymore, but an internal state.

From an application-driven point of view, these classifications can help to structure the system that is to do something useful, e.g., it makes sense using object-related SMCs for classifying objects, using intention-related SMCs for learning certain tasks, etc. From a more phenomenological point of view, however, the classification into concepts like modalities, objects and intentions seems to be imposed by the experimenter who looks at the system from outside in order to analyze it, whereas in an agent's situated perspective, these concepts might seem arbitrary and only a few among many possible entities of action-outcome relations. This is a classical instance of the frame-of-reference problem [Pfeifer and Bongard, 2007, p. 72ff.]. As an example, consider the terrain-specific information structures we found in Section 3.4 and which we attributed to the class of object-related SMCs by claiming that terrains, in a more abstract way, can be seen as objects (although terrains are not exactly what we immediately think of when we think about objects). From a sensorimotor- or information-based view, terrains and objects are environmental features that affect the SMCs of the legged robot a lot (and in special ways). We could thus also call them terrain-related SMCs in the case of the walking robot. For a flying robot, such different terrains might seem equal, as it rarely experiences the different friction properties of their surfaces, and importantly, makes no use of such knowledge for its own action selection. In order to reveal "clusters" of SMCs that are specific to the embodiment and situatedness of the agent under consideration, statistical learning and other unsupervised methods such as the information-theoretic ones we used seems beneficial.

In this respect, the SMC classes of Buhrmann et al. [2013] introduced in Section 2.2.2 provide a different spectrum of classification that avoids this problem. The four types of SMCs can be arranged along a dimension going from an external view towards an agent's perspective of the analysis. On the lowest level is the SM environment, which is merely limited by the physical constrains of the open-loop (brain-)body-environment system. No agent-specific intelligence is considered at all, the SM environment is rather spanning the space of possible sensorimotor trajectories that corresponds to all agents with similar sensors and actuators in similar environments. When closing the loop and adding the agent's internal dynamics, we traverse the more agent-centered SM habitat. The SM coordination and SM strategy incorporate task-related and normative dimensions that affect the agent's individual value system (and thus higher-level cognitive processes).

In [Buhrmann et al., 2013], the authors admit that practical application of their SMC descriptions in terms of dynamical systems is limited by the fact that all the relevant equations of a cognitive system are rarely known in practice. However, they argue that the four definitions can be applied to other methods and measures as well, such as information-theoretic or probabilistic ones. We demonstrated this first in Section 3.3 by viewing the information structure resulting from motor babbling as being the SM environment of the agent, and second in Section 4.2 by discussing how the SMCs learned with the probabilistic Markov model correspond to the SM habitat and when connected to a value system to the SM strategy of the agent.

6.3 Action sampling strategies

A recurrently upcoming theme in our studies of SMCs in artificial agents is the selection of actions in order to sample the sensorimotor space in an efficient way. All presented SMC learning methods, the information-theoretic one, the Markov model, the actor-critic design and the inverse kinematic model require some form of motor babbling, a random exploration in order to generate training samples from which the SMCs can be learned. This bears the problem of finding a strategy to efficiently sample relevant parts of the space. Uniformly sampling the whole space is usually infeasible for more complex agents with high-dimensional continuous sensorimotor spaces (even though biomechanical constraints already reduce the dimensionality drastically). Classically, ϵ-greedy policies are used that implement some kind of bootstrapping method. The ϵ defines the probability to select a random action, which can be drawn uniformly, but is often a normal distribution centered at the previously selected action. Otherwise, with probability $(1-\epsilon)$, the action is selected by the control method used (ESN-actor, inverse model, etc). The choice of ϵ represents the trade-off between exploring unknown parts of the space (random action) and exploiting the learned knowledge of the system. In some of our experiments we used a gradually decreasing ϵ, in order to start with completely random actions and let the control method take over control more and more over time. In fact, in those experiments, this was the only way to ensure the learning does not get stuck in the very beginning by always choosing trivial and useless actions. However, even with time-varying ϵ, such bootstrapping methods might take a long time before they visit a large enough region of the state-action space and fully unfold their capabilities.

As mentioned before, several approaches have been suggested, sometimes depending on the robotic platform used or the specific task that is to be achieved. For robotic arm reaching, the goal-babbling strategy has been developed by Rolf et al. [2011] and demonstrated to be successful on a multiple degrees of freedom robotic arm with pneumatic actuators [Hartmann et al., 2012]. Here, each training sample is assigned a weight indicating how much it influences the training (i.e., how much the current training parameters are changed). The weighting is larger, if the performed action moves the end-effector closer towards the target position and smaller otherwise. This way, a large amount of the sensorimotor space is ignored or down-graded and only the regions relevant for the task are learned. The method also provided a solution to the redundancy problem that multiple joint trajectories are able to achieve the desired task (also known as Bernstein's problem [Bernstein, 1967]). Other approaches deal with freezing and subsequent freeing of degrees of freedom (DoF). Here, several distal DoFs are locked in the beginning and only proximal ones are trained. Once training converged, the next DoF is freed and included in the training. This successive learning goes on until all DoFs are unlocked. Some authors even showed that alternating freezing and freeing is required for optimal performance [Berthouze and Lungarella, 2004].

Another very promising and more general strategy, which is independent of the platform, is the following. When the learning includes a prediction of a fitness or reward function, Gaussian process regression could be employed to predicts the fitness landscape from sensorimotor inputs. Gaussian processes have the advantage that they provide an uncertainty measure together with the prediction. This uncertainty can be used in an exploration-exploitation trade-off, by selecting actions from regions where the predicted fitness seems high, but has still high uncertainty (because not visited so far). This strategy has been demonstrated successful in machine learning in the context of optimal protein folding [Romero et al., 2013].

Machine Learning provides a further branch called *active learning*, which is also concerned with generation of training examples in an efficient way. The goal is to minimize the number of samples needed while achieving or maintaining a certain level of performance [Cohn et al., 1996]. Active learning has been used in the context of developmental robotics in [Baranes and Oudeyer, 2013].

Another interesting alternative to random exploration is provided by Martius et al. [2013] who maximize time-local predictive information using a gradient method. Since the control is deterministic here and has parameters that vary based on the information measure, the authors argue that this method implements a self-determined spontaneity (which they call exploration dynamics): "a self-directed search process, where the directions to explore are created from the dynamics of the system itself" [Martius et al., 2013, p. 12], rather than from a random number generator in the action-selection scheme.

Finally, our eye movement study from Section 5.2 provides the basis for a new bottom-up mechanism for visual attention that could be used in robotics to control camera or head movements of a robot. Such a control mechanism would trade off selecting movements to new parts of the scene, that foster exploration of the visual scenery versus selection of movements to interesting regions that contain relevant features worth investigating in more detail. This possibility is so far limited to visual exploration. A generalization of this concept to other modalities (e.g., arm movements or walking) could model the saccadic momentum as a momentum on the previous action. However, it is not clear how saliency could be encoded in a proprioceptive or inertia sensorimotor space.

6.4 Towards a formal framework for SMCT in AI

As we see from the amount of studies listed in Section 2.1, many methods are related to the sensorimotor approach without explicitly referring to SMCT. One major problem with SMCT is that to date no formal treatment of the ideas exist, which makes it difficult to operationalize it in robotics. Furthermore, important questions left open by the original publication of this theory [O'Regan and Noë, 2001] are (1.) which "structure" in sensory changes that follow movements should be considered and (2.) how this structure should be extracted and stored and (3.) how to be activated for action selection.

We proposed to answer the first three questions in terms of an information-theoretic analysis that quantifies the SMCs as the amount of information that is transferred between motor and sensory channels. We demonstrated how bodily properties as well as environmental structures can be identified with this approach. However, a clear directive how to utilize these for action selection and task solving behavior (fourth question) is still missing, although we point out some promising developments in that direction [Jung et al., 2011; Zahedi et al., 2013].

In our work we tackled the fourth question with a different approach. Using a reinforcement learning framework allowed us to formalize the task as a reward function and using appropriate tools from machine learning enabled us to learn and activate the SMCs that affect this reward. In a very similar way, Maye and Engel [2011, 2012a,b, 2013] set out to extract a clear definition from SMCT and suggest a treatment in terms of reinforcement learning, as it provides a universal framework that can be used in the context of the sensorimotor theory when states and actions are defined appropriately. Furthermore, reinforcement learning is a well studied theory with a comprehensive backup of mathematical descriptions to formalize the system and processes under investigation. It was indeed argued that reinforcement learning is a promising candidate to a general formal definition of artificial intelligence, because any formal problem can be encoded in a reward function [Schmidhuber, 2007].

As another attempt of a formal treatment of SMCT, the work by Buhrmann et al. [2013] is a step in a promising direction, in that they propose to formalize the structure of SMCs as a set of dynamical properties (e.g., attractors, bifurcations, metastable regions) in the sensorimotor space. With this dynamical systems account they answer the first three questions posed above. Their formalization allows for observational analyses of cognitive systems, a directly applicable design principle or suggestion for learning mechanisms or action selection strategy is, however, left open.

Nevertheless their distinction of the different SMC regimes contributes to the clarification of the SMCT approach.

More of such work is required in future in order to enhance practical applicability of SMCT to artificial systems and to promote it as a powerful theory for artificial intelligence and robotics.

6.5 Conclusions

The sensorimotor approach to cognition encompasses all the studies discussed in this thesis. Although these stand alone as single research papers with their own methodology and research contribution, they follow a common core idea of how learning, perception and cognition can be viewed in artificial intelligence. Namely as a pragmatic process where the action is of major importance for the emergence of intelligence. Studying artificial agents in terms of SMCT, a novel and promising sensorimotor theory within the pragmatic turn in cognitive science, follows some underlying principles which connect all the discussed approaches.

The most important being the rejection of symbolic representations of the world that are stored and recalled for performing computations on. Cognition can not merely be described by passive processing of external stimuli in order to derive world models. It can rather be seen as world making, which relies on exercise and mastery of the sensorimotor relations (or action-outcome relations on higher cognitive levels) through persistent interaction with the environment. Thus, grounding cognition in the sensorimotor knowledge that emerges from interaction with the environment is what is common to all the presented studies.

We discussed several possibilities of how this kind of sensorimotor knowledge can be captured and represented. Information theory and machine learning provide abundant methodology for learning and representation of such relations. However, a major commonality we identify among all those methods is the motor babbling approach, the self-initiated active generation of sensory stimulations from which SMCs are sampled. Artificial systems that implement SMCT are thus self-explorative systems that learn the sensorimotor relations incrementally from bottom-up. We showed how such a sensorimotor knowledge provides the basis from which concepts such as modalities or objects can be learned (see Chapter 3). This knowledge can then be exploited for control either by directly mapping the desired behavior onto the right actions (see inverse model learning in Section 5.1) or by deriving higher-level relations from them through providing external feedback in form of rewards (see Chapter 4 on SMCs in reinforcement learning).

Another commonality between our SMC approaches to artificial agents was to keep the amount of externally provided knowledge at a minimum level. We did not provide any meaning (e.g., in form of symbols) to the robots, but let them autonomously extract relevant and meaningful elements from the sensorimotor space. Being sufficient on lower levels, it was, however, necessary to provide external knowledge at higher levels, at least about the nature of the task the robot was performing (by specifying the reward function in reinforcement learning, or the target locations in the reaching study).

In conclusion, we showed that by adopting the pragmatic approach to cognition in artificial intelligence, we can achieve some of the requirements posed in the introduction. SMC learning provides us the kind of generalization, robustness and adaptivity that is necessary for real-world applications. We furthermore put SMCT a critical step further by extending the notion of SMCs beyond low-level relations among sensory and motor signals. By introducing intention-related SMCs and traversing the domain of SM strategy, we contributed to an extension to higher levels of cognition. Further work in this direction, together with a more elaborated formal treatment, might bring SMCT towards a full-fledged theory of cognition.

References

C. Alessandro, I. Delis, F. Nori, S. Panzeri, and B. Berret. Muscle synergies in neuroscience and robotics: from input-space to task-space perspectives. *Frontiers in computational neuroscience*, 7 (April):43, Jan. 2013. ISSN 1662-5188. doi: 10.3389/fncom.2013.00043.

N. Ay and D. Polani. Information flows in causal networks. *Advances in complex systems*, 11(1): 17–41, 2008. doi: 10.1142/S0219525908001465.

N. Ay, N. Bertschinger, R. Der, F. Güttler, and E. Olbrich. Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B*, 63(3):329–339, Apr. 2008. ISSN 1434-6028. doi: 10.1140/epjb/e2008-00175-0.

C. Bandt and B. Pompe. Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical Review Letters*, 88(17):1–4, Apr. 2002. ISSN 0031-9007. doi: 10.1103/PhysRevLett.88.174102.

A. Baranes and P.-Y. Oudeyer. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, 61(1):49–73, 2012. ISSN 09218890. doi: 10.1016/j.robot.2012.05.008.

A. Baranes and P.-Y. Oudeyer. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems*, 61(1):49–73, Jan. 2013. ISSN 09218890. doi: 10.1016/j.robot.2012.05.008.

M. Baumgartner. *Heuristic Dynamic Programming with Echo State Networks*. Master thesis, University of Zurich, 2013.

P. M. Bays and D. M. Wolpert. Computational principles of sensorimotor control that minimize uncertainty and variability. *The Journal of physiology*, 578(Pt 2):387–96, Jan. 2007. ISSN 0022-3751. doi: 10.1113/jphysiol.2006.120121.

R. Beer. A dynamical systems perspective on agent-environment interaction. *Artificial intelligence*, 72:173–215, 1995. doi: 10.1016/0004-3702(94)00005-L.

R. D. Beer. The Dynamics of Brain– Body–Environment Systems: A Status Report. In A. Gomila and C. P., editors, *Handbook of Cognitive Science: An Embodied Approach*, pages 99–120. Elsevier, 2008.

M. Begum and F. Karray. Visual Attention for Robotic Cognition: A Survey. *IEEE Transactions on Autonomous Mental Development*, 3(1):92–105, Mar. 2011. ISSN 1943-0604. doi: 10.1109/TAMD.2010.2096505.

N. A. Bernstein. *The Coordination and Regulation of Movements*. Oxford, New York, Pergamon Press, 1st edition, 1967.

L. Berthouze and M. Lungarella. Motor Skill Acquisition Under Environmental Perturbations: On the Necessity of Alternate Freezing and Freeing of Degrees of Freedom. *Adaptive Behavior*, 12(1):47–64, Mar. 2004. ISSN 10597123. doi: 10.1177/105971230401200104.

S. Bovet and R. Pfeifer. Emergence of Delayed Reward Learning from Sensorimotor Coordination. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2272–2277. Ieee, 2005. doi: 10.1109/IROS.2005.1545085.

V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA, 1986. ISBN 9780262521123.

R. A. Brooks. A robust layered control system for a mobile robot. *Robotics and Automation, IEEE Journal of*, 2(1):14–23, 1986. doi: 10.1109/JRA.1986.1087032.

R. A. Brooks. Elephants don't play chess. *Robotics and Autonomous Systems*, 6:3–15, 1990. doi: 10.1016/S0921-8890(05)80025-9.

R. A. Brooks. Intelligence without representation. *Artificial Intelligence Journal*, 47:139–159, Oct. 1991. ISSN 1364-503X. doi: 10.1016/0004-3702(91)90053-M.

T. Buhrmann, E. A. Di Paolo, and X. Barandiaran. A dynamical systems account of sensorimotor contingencies. *Frontiers in Psychology*, 4(May):285, Jan. 2013. ISSN 1664-1078. doi: 10.3389/fpsyg.2013.00285.

A. Clark. *Being there: Putting Brain, Body, and World Together Again*. 1997. ISBN 9780262531566.

A. Clark and D. Chalmers. The Extended Mind. *Analysis*, 58(1):7–19, 1998.

D. Cohn, Z. Ghahramani, and M. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edition, 2006. ISBN 9780471241959. doi: 10.1002/047174882X.

Y. Demiris and A. Dearden. From motor babbling to hierarchical learning by imitation: a robot developmental pathway. In L. Berthouze, F. Kaplan, H. Kozima, H. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov, and C. Balkenius, editors, *Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, pages 31–37, Lund, 2005. ISBN 9197474142.

R. Der and G. Martius. From motor babbling to purposive actions: Emerging self-exploration in a dynamical systems approach to early robot development. In S. Nolfi, G. Baldassarre, R. Calabretta, J. C. T. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, and D. Parisi, editors, *From Animals to Animats 9: 9th International Conference on Simulation of Adaptive Behavior, SAB 2006*, pages 406–421, Rome, Italy, 2006. Springer Berlin Heidelberg. doi: 10.1007/11840541_34.

J. Dewey. The reflex arc concept in psychology. *Psychological review*, 3(4):357–370, 1896. doi: 10.1037/h0070405.

E. A. Di Paolo, M. Rohde, and H. De Jaegher. Horizons for the Enactive Mind : Values , Social Interaction, and Play. In J. Stewart, O. Gapenne, and E. A. Di Paolo, editors, *Enaction: Toward a New Paradigm for Cognitive Science*, number April, chapter 2, pages 33–87. The MIT Press, 2010. ISBN 9780262014601.

E. A. Di Paolo, X. E. Barandiaran, M. Beaton, and T. Buhrmann. Learning to perceive in the sensorimotor approach: Piaget's theory of equilibration interpreted dynamically. *Frontiers in Human Neuroscience*, 8(July):1–16, July 2014. ISSN 1662-5161. doi: 10.3389/fnhum.2014.00551.

A. K. Engel. Directive minds: how dynamics shapes cognition. In J. Stewart, O. Gapenne, and E. A. Di Paolo, editors, *Enaction: Toward a New Paradigm for Cognitive Science*, chapter 8, pages 219–243. The MIT Press, 2010. ISBN 9780262014601.

A. K. Engel, A. Maye, M. Kurthen, and P. König. Where's the action? The pragmatic turn in cognitive science. *Trends in cognitive sciences*, 17(5):202–9, May 2013. ISSN 1879-307X. doi: 10.1016/j.tics.2013.03.006.

eSMCs. Consortium of the EU-FP-7 Project "Extending Sensorimotor Contingencies to Cognition" (eSMCs), 2010. FP7-ICT-270212, http://esmcs.eu.

J. Feldman and D. Ballard. Connectionist models and their properties. *Cognitive science*, 6(3): 205–254, 1982. doi: 10.1207/s15516709cog0603_1.

P. Fine, E. Di Paolo, and E. Izquierdo. Adapting to your body. In F. A. e. Costa, L. M. Rocha, E. Costa, I. Harvey, and A. Coutinho, editors, *Advances in Artificial Life*, pages 203–212, Lisbon, Portugal, 2007. Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-540-74913-4_21.

J. A. Fodor. *The Language of Thought*. Harvard University Press, 1975. ISBN 0-674-51030-5.

C. Hartmann, J. Boedecker, O. Obst, S. Ikemoto, and M. Asada. Real-Time Inverse Dynamics Learning for Musculoskeletal Robots based on Echo State Gaussian Process Regression. In N. Roy, P. Newman, and S. Srinivasa, editors, *Proceedings of Robotics: Science and Systems VIII*, Sydney, Australia, 2012. MIT Press.

J. Haugeland. *Artificial Intelligence: The Very Idea*. MIT Press, Cambridge, MA, 1985. ISBN 0-262-08153-9.

H. Hauser, A. J. Ijspeert, R. M. Füchslin, R. Pfeifer, and W. Maass. Towards a theoretical foundation for morphological computation with compliant bodies. *Biological cybernetics*, (2011):355–370, Jan. 2012. ISSN 1432-0770. doi: 10.1007/s00422-012-0471-0.

K. Hlavackovaschindler, M. Palus, M. Vejmelka, and J. Bhattacharya. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, 441(1):1–46, Mar. 2007. ISSN 03701573. doi: 10.1016/j.physrep.2006.12.004.

M. Hoffmann, H. G. Marques, A. H. Arieta, H. Sumioka, M. Lungarella, and R. Pfeifer. Body schema in robotics: A review. *IEEE transactions on autonomous mental development*, 2(4):304–324, 2010.

M. Hoffmann, N. M. Schmidt, R. Pfeifer, A. K. Engel, and A. Maye. Using sensorimotor contingencies for terrain discrimination and adaptive walking behavior in the quadruped robot puppy. In T. Ziemke, C. Balkenius, and J. Hallam, editors, *From Animals to Animats 12*, volume 7426, pages 54–64, Odense, Denmark, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-33092-6. doi: 10.1007/978-3-642-33093-3_6.

V. Hogman, M. Bjorkman, and D. Kragic. Interactive object classification using sensorimotor contingencies. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2799–2805, Nov. 2013. doi: 10.1109/IROS.2013.6696752.

S. Horst. The Computational Theory of Mind. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 201 edition, 2011.

H. Imamizu and M. Kawato. Cerebellar internal models: implications for the dexterous use of tools. *Cerebellum*, 11(2):325–35, June 2012. ISSN 1473-4230. doi: 10.1007/s12311-010-0241-2.

L. Itti and C. Koch. Computational modelling of visual attention. *Nature reviews. Neuroscience*, 2 (3):194–203, Mar. 2001.

H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note'. Technical report, 2001.

T. Jung, D. Polani, and P. Stone. Empowerment for continuous agent–environment systems. *Adaptive Behavior*, 19(1):16–39, Jan. 2011. ISSN 1059-7123. doi: 10.1177/1059712310392389.

R. Klein and W. MacInnes. Inhibition of return is a foraging fascilitator in visual search. *Psychological science*, 10(4):346–352, 1999.

A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment : A Universal Agent-Centric Measure of Control. In *The 2005 IEEE Congress on Evolutionary Computation, 2005.*, pages 128–135, Edinburgh, Scotland, 2005a. IEEE. doi: 10.1109/CEC.2005.1554676.

A. S. Klyubin, D. Polani, and C. L. Nehaniv. All Else Being Equal Be Empowered. In *Advances in Artificial Life, European Conference on Artificial Life (ECAL)*, pages 744–753, Canterbury, UK, 2005b. Springer Berlin Heidelberg. doi: 10.1007/11553090_75.

J. Kober and J. Peters. Policy search for motor primitives in robotics. *Machine Learning*, 84(1-2): 171–203, Nov. 2010. ISSN 0885-6125. doi: 10.1007/s10994-010-5223-6.

G. Kootstra, N. Wilming, N. M. Schmidt, M. Djurfeldt, D. Kragic, and P. König. Learning and Adaptation of Sensorimotor Contingencies: Prism-Adaptation, a Case Study. In T. Ziemke, C. Balkenius, and J. Hallam, editors, *From Animals to Animats 12*, pages 341–350, Odense, Denmark, 2012. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-33093-3_34.

P. Koprinkova-Hristova, M. Oubbati, and G. Palm. Adaptive Critic Design with Echo State Network. *2010 IEEE International Conference on Systems, Man and Cybernetics*, pages 1010–1015, Oct. 2010. doi: 10.1109/ICSMC.2010.5641744.

A. S. Kornheiser. Adaptation to laterally displaced vision: a review. *Psychological bulletin*, 83(5): 783–816, 1976. doi: 10.1037/0033-2909.83.5.783.

G. Lakoff and M. Johnsen. *Metaphors We Live By*. The University of Chicago Press, 1980. ISBN 9780226468013.

M. Lungarella and O. Sporns. Mapping information flow in sensorimotor networks. *PLoS computational biology*, 2(10):e144, Oct. 2006. ISSN 1553-7358. doi: 10.1371/journal.pcbi.0020144.

M. Lungarella, T. Pegors, D. Bulwinkle, and O. Sporns. Methods for Quantifying the Informational Structure of Sensory and Motor Data. *Neuroinformatics*, 3(3):243–262, 2005. doi: 10.1385/NI:3:3:243.

H. G. Marques, A. Bharadwaj, and F. Iida. From spontaneous motor activity to coordinated behaviour: a developmental model. *PLoS computational biology*, 10(7):e1003653, July 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003653.

G. Martius, R. Der, and N. Ay. Information driven self-organization of complex robotic behaviors. *PloS one*, 8(5):e63400, Jan. 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0063400.

A. Maye and A. K. Engel. A discrete computational model of sensorimotor contingencies for object perception and control of behavior. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3810–3815, 2011. doi: 10.1109/ICRA.2011.5979919.

A. Maye and A. K. Engel. Time Scales of Sensorimotor Contingencies. In H. Zhang, A. Hussain, D. Liu, and Z. Wang, editors, *Advances in Brain Inspired Cognitive Systems*, pages 240–249, Shenyang, Liaoning, China, 2012a. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-31561-9_27.

A. Maye and A. K. Engel. Using Sensorimotor Contingencies for Prediction and Action Planning. In T. Ziemke, C. Balkenius, and J. Hallam, editors, *From Animals to Animats 12*, volume 7426, pages 106–116, Odense, Denmark, 2012b. Springer Berlin Heidelberg. doi: 10.1007/978-3-642-33093-3_11.

a. Maye and a. K. Engel. Extending sensorimotor contingency theory: prediction, planning, and action generation. *Adaptive Behavior*, 21(6):423–436, Aug. 2013. ISSN 1059-7123. doi: 10.1177/1059712313497975.

J. McClelland, D. Rumelhart, and G. Hinton. The Appeal of Parallel Distributed Processing. In D. E. Rumelhart, J. L. McClelland, and PDP Research Group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*, chapter 1, pages 3–44. MIT Press, Cambridge, MA, 1986. ISBN 9780262181204.

T. McGeer. Passive Dynamic Walking. *The International Journal of Robotics Research*, 9(2):62–82, 1990. doi: 10.1177/027836499000900206.

A. Meltzoff and M. Moore. Explaining facial imitation: A theoretical model. *Early development and parenting*, 6(3-4):179–192, 1997. doi: 10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R.

M. Merleau-Ponty. *The Structure of Behavior*. Beacon Press, Boston, 1963. ISBN 0-8070-2987-4.

M. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA, expanded edition, 1987. ISBN 9780262631112.

R. Möller and W. Schenck. Bootstrapping cognition from behavior-a computerized thought experiment. *Cognitive science*, 32(3):504–42, Apr. 2008. ISSN 0364-0213. doi: 10.1080/03640210802035241.

H. Murao, H. Tamaki, and S. Kitamura. Walking pattern acquisition for quadruped robot by using modular reinforcement learning. *2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace*, 3:1402–1405, 2001. doi: 10.1109/ICSMC.2001.973478.

K. Nakajima, T. Li, H. Sumioka, M. Cianchetti, and R. Pfeifer. Information Theoretic Analysis on a Soft Robotic Arm Inspired by the Octopus. In *Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics, ROBIO 2011*, volume 1, pages 110–117, Phuket, Thailand, 2011. IEEE. ISBN 9781457721373. doi: 10.1109/ROBIO.2011.6181271.

K. Nakajima, N. M. Schmidt, and R. Pfeifer. Measuring information transfer in a soft robotic arm. *Bioinspiration & Biomimetics*, 10(3):035007, 2015. doi: 10.1088/1748-3190/10/3/035007.

A. Newell and H. Simon. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126, 1976.

D. Nguyen-Tuong and J. Peters. Model learning for robot control: a survey. *Cognitive processing*, 2011.

D. Nguyen-Tuong, M. Seeger, and J. Peters. Model Learning with Local Gaussian Process Regression. *Advanced Robotics*, 23(15):2015–2034, Jan. 2009. ISSN 0169-1864. doi: 10.1163/016918609X12529286896877.

J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *The Behavioral and brain sciences*, 24(5):939–73; discussion 973–1031, Oct. 2001. ISSN 0140-525X.

M. Oubbati, J. Uhlemann, and G. Palm. Adaptive Learning in Continuous Environment Using Actor-Critic Design and Echo-State Networks. *From Animals to Animats 12*, pages 320–329, 2012.

P.-Y. Oudeyer and F. Kaplan. What is Intrinsic Motivation? A Typology of Computational Approaches. *Frontiers in neurorobotics*, 1(November):6, Jan. 2007. ISSN 1662-5218. doi: 10.3389/neuro.12.006.2007.

P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner. Intrinsic motivation systems for autonomous mental development. *IEEE Transactions on Evolutionary Computation*, 11(2):265–286, 2007. doi: 10.1109/TEVC.2006.890271.

R. Pfeifer and J. Bongard. *How the Body Shapes the Way We Think: A New View of Intelligence*. Bradford Books, MIT Press, 2007. ISBN 9780262162395.

R. Pfeifer and C. Scheier. Sensory-motor coordination: The metaphor and beyond. *Robotics and Autonomous Systems*, 20(2-4):157–178, 1997. doi: 10.1016/S0921-8890(97)80707-5.

R. Pfeifer and C. Scheier. *Understanding Intelligence*. MIT Press, Cambridge, MA, 1999. ISBN 0262161818.

R. Pfeifer, M. Lungarella, O. Sporns, and Y. Kuniyoshi. On the Information Theoretic Implications of Embodiment – Principles and Methods. In M. Lungarella, F. Iida, J. Bongard, and R. Pfeifer, editors, *50 Years of AI*, pages 76–86. Springer-Verlag Berlin Heidelberg, 2007. ISBN 978-3-540-77296-5. doi: 10.1007/978-3-540-77296-5_8.

R. Pfeifer, M. Lungarella, and F. Iida. The challenges ahead for bio-inspired 'soft' robotics. *Communications of the ACM*, 55(11):76–87, 2012. doi: 10.1145/2366316.2366335.

J. Piaget. *The Origins of Intelligence in Children*. International Universities Press, New York, translated from French original "La naissance de l'intelligence chez l'enfant" (1936), 1952.

D. Polani. Information: Currency of life? *HFSP Journal*, 3(5):307–316, 2009. doi: 10.2976/1.3171566.

D. Polani, O. Sporns, and M. Lungarella. How Information and Embodiment Shape Intelligent Information Processing. In M. Lungarella, F. Iida, J. Bongard, and R. Pfeifer, editors, *50 Years of AI*, pages 99–111. Springer-Verlag Berlin Heidelberg, 2007. ISBN 978-3-540-77296-5. doi: 10.1007/978-3-540-77296-5_10.

B. Pompe and J. Runge. Momentary information transfer as a coupling measure of time series. *Physical Review E*, 83:051122, May 2011. ISSN 1539-3755. doi: 10.1103/PhysRevE.83.051122.

M. Posner and Y. Cohen. Components of Visual Orienting. In H. Bouma and D. Bouwhuis, editors, *Attention and performance X: Control of language processes*, volume 32, pages 531–556. Hillsdale, NJ: Erlbaum, 1984.

M. Prokopenko, V. Gerasimov, and I. Tanev. Evolving spatiotemporal coordination in a modular robotic system. In S. Nolfi, G. Baldassarre, R. Calabretta, J. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, and D. Parisi, editors, *From Animals to Animats 9: Proceedings of the 9th International Conference on Simulation of Adaptive Behavior, (SAB2006)*, pages 558–569. Springer Berlin Heidelberg, Rome, Italy, 2006. ISBN 978-3-540-38608-7. doi: 10.1007/11840541_46.

H. Putnam. *Mind, Language and Reality. Philosophical Papers: Volume 2*. Cambridge University Press, 1975. ISBN 9780521295512.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*, volume 14. MIT Press, Cambridge, MA, 2nd edition, Apr. 2006. ISBN 026218253X.

G. M. Redding and B. Wallace. Generalization of prism adaptation. *Journal of experimental psychology. Human perception and performance*, 32(4):1006–22, Aug. 2006. ISSN 0096-1523. doi: 10.1037/0096-1523.32.4.1006.

M. Rolf, J. J. Steil, and M. Gienger. Online Goal Babbling for rapid bootstrapping of inverse models in high dimensions. In *2011 IEEE International Conference on Development and Learning (ICDL)*, pages 1–8. Ieee, Aug. 2011. ISBN 978-1-61284-989-8. doi: 10.1109/DEVLRN.2011.6037368.

P. a. Romero, A. Krause, and F. H. Arnold. Navigating the protein fitness landscape with Gaussian processes. *Proceedings of the National Academy of Sciences of the United States of America*, 110:E193–201, 2013. ISSN 1091-6490. doi: 10.1073/pnas.1215251110.

J. Ruesch, M. Lopes, A. Bernardino, J. Hornstein, J. Santos-Victor, and R. Pfeifer. Multimodal saliency-based bottom-up attention a framework for the humanoid robot iCub. In *IEEE International Conference on Robotics and Automation*, pages 962–967, Pasadena, CA, USA, May 2008. Ieee. ISBN 978-1-4244-1646-2. doi: 10.1109/ROBOT.2008.4543329.

C. Salge, C. Glackin, and D. Polani. Empowerment — An Introduction. In M. Prokopenko, editor, *Guided Self-Organization: Inception*, pages 67–114. Springer Berlin Heidelberg, 2014. ISBN 978-3-642-53734-9. doi: 10.1007/978-3-642-53734-9_4.

J. Schmidhuber. Curious model-building control systems. In *IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 1458–1463, Singapore, 1991. Ieee. ISBN 0-7803-0227-3. doi: 10.1109/IJCNN.1991.170605.

J. Schmidhuber. 2006 : Celebrating 75 Years of AI - History and Outlook : The Next 25 Years. In M. Lungarella, F. Iida, R. Pfeifer, and J. Bongard, editors, *50 Years of Artificial intelligence*, pages 29–41. Springer-Verlag Berlin Heidelberg, 2007. ISBN 978-3-540-77295-8. doi: 10.1007/978-3-540-77296-5_4.

N. M. Schmidt, M. Hoffmann, and K. Nakajima. Information flow in a quadruped running robot quantified by transfer entropy. In *5th International Conference on Cognitive Systems (CogSys 2012)*, page 112, Vienna, Austria, 2012.

N. M. Schmidt, M. Hoffmann, K. Nakajima, and R. Pfeifer. Bootstrapping Perception Using Information Theory: Case Studies in a Quadruped Robot Running on Different Grounds. *Advances in Complex Systems*, 16(2n03):1250078, July 2013. ISSN 0219-5259. doi: 10.1142/S0219525912500786.

N. M. Schmidt, M. Baumgartner, and R. Pfeifer. Actor-Critic Design using Echo State Networks in a Simulated Quadruped Robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2224–2229, Chicago, IL, USA, 2014. Ieee. doi: 10.1109/IROS.2014.6942862.

T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464, July 2000. ISSN 1079-7114. doi: 10.1103/PhysRevLett.85.461.

O. Sigaud, C. Salaün, and V. Padois. On-line regression algorithms for learning mechanical models of robots: A survey. *Robotics and Autonomous Systems*, 59(12):1115–1129, Dec. 2011. ISSN 09218890. doi: 10.1016/j.robot.2011.07.006.

T. J. Smith and J. M. Henderson. Facilitation of return during scene viewing. *Visual Cognition*, 17 (6-7):1083–1108, Aug. 2009. ISSN 1350-6285. doi: 10.1080/13506280802678557.

O. Sporns and M. Lungarella. Evolving coordinated behavior by maximizing information structure. In L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone, and A. Vespignani, editors, *Artificial life X: proceedings of the tenth international conference on the simulation and synthesis of living systems.*, pages 323–329, Bloomington, IN, USA, 2006. The MIT Press.

L. Steels. The artificial life roots of artificial intelligence. *Artificial Life*, 1(1-2):75–110, 1993. doi: 10.1162/artl.1993.1.1_2.75.

R. Sutton and A. Barto. *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA, 1998 A Bradford Book, 1998.

F. J. Varela, E. Thompson, and E. Rosch. *The Embodied Mind: Cognitive Science and Human Experience*, volume 1992. MIT Press, Cambridge, MA, 1991. ISBN 9780262220422.

W. G. Walter. *The Living Brain*. Norton, New York, 1953.

M. Wheeler. Embodied Cognition and the Extended Mind. In J. Garvey, editor, *The Continuum Companion to Philosophy of Mind*, chapter 12, pages 220–238. Continuum, 2011. ISBN 9780826431882.

M. Wibral, R. Vicente, and M. Lindner. Transfer Entropy in Neuroscience. In M. Wibral, R. Vicente, and J. T. Lizier, editors, *Directed Information Measures in Neuroscience*, pages 3–36. Springer Berlin Heidelberg, 1st edition, 2014. ISBN 978-3-642-54473-6. doi: 10.1007/978-3-642-54474-3_1.

P. L. Williams and R. D. Beer. Information dynamics of evolved agents. In S. Doncieux, B. Girard, A. Guillot, J. Hallam, J.-A. Meyer, and J.-B. Mouret, editors, *From Animals to Animats 11: Proceedings of the 11th International Conference on Simulation of Adaptive Behavior*, pages 38–49. Springer, 2010.

P. L. Williams and R. D. Beer. Generalized measures of information transfer. *arXiv preprint arXiv:1102.1507*, pages 1–6, 2011.

N. Wilming, T. Betz, T. Kietzmann, and P. König. Measures and Limits of Models of Fixation Selection. *PloS one*, 6(9):e24038, 2011. doi: 10.1371/Citation.

N. Wilming, S. Harst, N. M. Schmidt, and P. König. Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy. *PLoS computational biology*, 9(1):e1002871, Jan. 2013. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002871.

D. M. Wolpert, R. C. Miall, and M. Kawato. Internal models in the cerebellum. *Trends in cognitive sciences*, 2(9):338–47, Sept. 1998. ISSN 1364-6613.

K. Zahedi, G. Martius, and N. Ay. Linear combination of one-step predictive information with an external reward in an episodic policy gradient setting : a critical analysis. *Frontiers in Psychology*, 4(801):1–13, 2013. doi: 10.3389/fpsyg.2013.00801.

Appendix A

Bootstrapping Perception using Information Theory: Case Studies in a Quadruped Robot Running on Different Grounds

Electronic version of an article published as:

Schmidt, N. M., Hoffmann, M., Nakajima, K. and Pfeifer, R. (2013). "Bootstrapping perception using information theory: Case studies in a quadruped robot running on different grounds." *Advances in Complex Systems*, 16(2n03):1250078, 2013, doi: 10.1142/S0219525912500786

BOOTSTRAPPING PERCEPTION USING INFORMATION THEORY: CASE STUDIES IN A QUADRUPED ROBOT RUNNING ON DIFFERENT GROUNDS

NICO M. SCHMIDT

Artificial Intelligence Laboratory, Department of Informatics, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland
nico.schmidt@uzh.ch

MATEJ HOFFMANN
hoffmann@ifi.uzh.ch

KOHEI NAKAJIMA
nakajima@ifi.uzh.ch

ROLF PFEIFER
pfeifer@ifi.uzh.ch

Received 21 February 2012
Revised 23 May 2012
Accepted 29 May 2012

Animals and humans engage in an enormous variety of behaviors which are orchestrated through a complex interaction of physical and informational processes: the physical interaction of the bodies with the environment is intimately coupled with informational processes in the animal's brain. A crucial step toward the mastery of all these behaviors seems to be to understand the flows of information in the sensorimotor networks. In this study, we have performed a quantitative analysis in an artificial agent - a running quadruped robot with multiple sensory modalities - using tools from information theory (transfer entropy). Starting from very little prior knowledge, through systematic variation of control signals and environment, we show how the agent can discover the structure of its sensorimotor space, identify proprioceptive and exteroceptive sensory modalities, and acquire a primitive body schema. In summary, we show how the analysis of directed information flows in an agent's sensorimotor networks can be used to bootstrap its perception and development.

Keywords: information theory; transfer entropy; perception; developmental robotics; sensorimotor contingencies; body schema;

1. Introduction

Animals are constantly being confronted with a massive multidimensional flow of information that is sampled by their receptors and, after some preprocessing, relayed

to their brains. This information has to be processed for the animal to be able to take the right decisions and execute the actions that maximize its chances of survival. In addition, the organism and the environment are dynamically and reciprocally coupled and so are the sensory and motor signals. It is the sensorimotor networks (as opposed to purely sensory information) and the dynamic patterns that exist in them that provide the basis for further processing. Cognition is then best viewed as emerging from this dynamic sensorimotor coupling (e.g. [39, 26]).

The view just described holds for natural and artificial agents (i.e. animals and robots) alike. Robots - if they are to autonomously succeed in the real world - also need to extract the relevant information about their interaction with the environment. In order to understand the nature of the processing that is responsible for cognition, the prerequisite seems to be to quantify and analyze the structure of the information flow in these sensorimotor networks. The tools of information theory such as entropy, mutual information, integration, complexity and transfer entropy have proven useful in this respect. They have been applied to inspect information flows inside the brain (e.g. [8, 37, 11, 9]), as well as in the data collected from robots. Lungarella and Sporns [20] have conducted studies on robots that illustrate the effect of individual components of the sensorimotor loop on the information structure. In particular, they showed how a given sensorimotor coordinated behavior (such as foveation) can increase the information content that reaches a given sensor (an artificial retina). Manipulating the sensor morphology (log-polar transformation in this case) showed similar effects. Williams and Beer [41] conducted an information-theoretic analysis of a simple agent engaged in a categorization task. Nakajima et. al. [22] showed how directed information flow, measured by symbolic transfer entropy can help to characterize the force-propagation in an artificial octopus arm.

The bulk of the work described so far was adopting a largely descriptive perspective - given a behaving system (a brain, or a complete agent with sensors and actuators), the information flow and structure was analyzed. We have argued above that this is a key step to understand the behavior of the system. However, an alternative perspective is to look at the world through the eyes of the agent itself. Imagine an animal or robot has just been "born". Using its actuators, it can interact with the world, generating sensory stimulation. Without prior knowledge of its body, sensory apparatus, and the surrounding environment, how can it make sense of the sensorimotor signals it is experiencing? As the agent interacts with the environment, it will experience some patterns (regularities, contingencies) much more often than others - this is given by the agent's embodiment, the morphology and material properties of its body and the placement of its sensors ([14] provide an overview of case studies illustrating these effects). Remembering or representing those regularities will be useful to the agent. But where should the agent start? We think that it should start at the very basis: it should first learn the extent of its body, the things it can influence and what lies beyond its control and should be attributed to the environment.

Such a process has been observed in infants who spend substantial time in

their early months observing and touching themselves [33]. Through this process of babbling, intermodal redundancies, temporal contingencies, and spatial congruences are picked up. In such a process, the infant forms a model of its body (a body image or body schema, see e.g. [4, 3, 21]). A developmental approach can also be applied to robots [19, 40, 27]. To identify its own body and learn about its contingencies is a natural candidate to start the autonomous development in an artificial agent (see [13] for a review on self-models and their acquisition in robots). Several studies along those lines have been conducted: typically, they involve an upper torso humanoid robot that is observing the space in front of it with a camera. The goal is to identify the parts of the visual scene that belong to its body (its arms, for instance). Different assumptions can be employed: the robot is static and environment is varied [43], or, on the contrary, temporal contingency is exploited by the robot - the robot learns to recognize its body parts by moving them [7, 23, 10]. Some researchers attempt to start with even less prior knowledge: Olsson et al. [24, 25] and Philipona et al. [30] study cases, where the agent is confronted with raw uninterpreted sensory signals only. There is no preprocessing, no knowledge of geometry and the agent does not even know which signals come from which modality. In [30] a simple simulated agent learns to make the distinction between body and environment by observing over which part of the sensory channels it has complete control. Olsson et al. [24, 25] have collected data from a real robot and showed that using an information metric as distances between the sensory and motor channels, the robot is able to reveal the (mainly spatial) relationships from its morphology (eg. arrangement of camera pixels).

Our work is very much in line with the approach of Philipona et al. [30] and Olsson et al. [24, 25]. For our study, we have chosen a running quadruped robot with the following modalities: four motor signals, eight angular position sensors (4 on active and 4 on passive joints), 4 pressure sensors on the robot's feet, a 3-axis accelerometer and 3-axis gyroscope. The control signal and the environment (five different ground materials) are systematically varied and the sensorimotor data is collected. The information flows are analyzed using transfer entropy and the effect of the different conditions is investigated. Then, we adopt the perspective of the autonomous agent and show how the agent can use the information flows to: (1) derive a primitive body schema and infer the controllability of different sensory variables; (2) discriminate different environments; (3) discover the structure of its sensorimotor space (identify proprioceptive and exteroceptive modalities, group different modalities and extract topological relations); and (4) interpret the quantity of information flow to assess the utility of different sensory channels and its overall performance.

This article is structured as follows. In Sec. 2, we will first introduce the information theoretic methods used and the experimental setup. In Sec. 3, we report on the results of the experiments. A brief section desribing the robot's behavior from an observer perspective is followed by a detailed analysis of the information

flows under different conditions and their implications for the robot's autonomous development and perception. The paper is closed by a discussion, followed by a final conclusion and suggestions on future work.

2. Materials and Methods

In this section, we describe the information theoretic measures used, our experimental setup, and explain in detail how we analyzed the data in this paper.

2.1. *Information Theoretic Measures: The Transfer Entropy*

We use the term information in the Shannon sense, that is, to quantify statistical patterns in observed variables. Thus, the measures presented here are based on Shannon entropy [2]. Given a time series x_t from the system X, entropy $H(X)$ provides a measure of the average uncertainty, or information, calculated from the probability distribution $p(x_t)$ according to:

$$H(X) = -\sum_{x_t} p(x_t) \log p(x_t). \tag{1}$$

The association between two time series is often expressed as their mutual information

$$I(X;Y) = \sum_{x_t} \sum_{y_t} p(x_t, y_t) \log \frac{p(x_t, y_t)}{p(x_t)p(y_t)}. \tag{2}$$

which expresses the deviation from the assumption that both are independent from each other. However, mutual information also contains information that is shared by X and Y due to a common history and it is invariant under exchange of the two variables. As we were interested in characterizing the directed information flow between the time series, we used *transfer entropy* [34], which provides this directionality and removes the shared information. Transfer entropy was introduced to measure the magnitude and the direction of information flow from one element to another and has been used to analyze information flows in real time series data from neuroscience [9, 11], robotics [38, 22], and many other fields. Given two time series X and Y, the transfer entropy TE essentially quantifies the deviation from the generalized Markov property $p(x_{t+1}|x_{t-\tau}) = p(x_{t+1}|x_{t-\tau}, y_{t-\tau})$ [34]. If the deviation is small, then $Y_{t-\tau}$ can be assumed to have little relevance on the transition from $X_{t-\tau}$ to X_{t+1}. If the deviation is large, however, then $Y_{t-\tau}$ adds information about the transition of $X_{t-\tau}$ and the generalized Markov property is not valid. The deviation from this assumption can, similar as in the mutual information, be expressed as a specific version of the Kullback-Leibler divergence:

$$TE_\tau(Y \to X) = \sum_{x_{t+1}} \sum_{x_{t-\tau}} \sum_{y_{t-\tau}} p(x_{t+1}, x_{t-\tau}, y_{t-\tau}) \log \frac{p(x_{t+1}|x_{t-\tau}, y_{t-\tau})}{p(x_{t+1}|x_{t-\tau})} \tag{3}$$

where the sums are over all possible states, t is the current time step and $\tau \in \mathbb{N}_0$ indicates the time lag of the transition.

In other words, TE measures how well we can predict the transition of the system X by knowing the system Y, beyond the degree to which X already disambiguates its own future. Transfer entropy is non-negative, any information transfer between the two variables resulting in $TE \geq 0$.

As proposed by Williams & Beer [42], the transfer entropy can be decomposed in two different *kinds* of information transfer, the *state-dependent transfer entropy* ($SDTE$) and the *state-independent transfer entropy* ($SITE$). The former characterizes the information transfer that is caused by the *synergy* of both variables in predicting the transition from $X_{t-\tau}$ to X_{t+1}, so it not only depends on $Y_{t-\tau}$, but also on the state of $X_{t-\tau}$. The latter kind of information transfer is the *unique information* that $Y_{t-\tau}$ yields about X_{t+1} and is completely independent from $X_{t-\tau}$. Moreover, in control theoretic terms the $SITE$ expresses the open-loop controllability of a variable X by its controller Y, while the $SDTE$ expresses the Y's closed-loop controllability of X [42].

The state-dependent and state-independent transfer entropy are defined as

$$SITE_\tau(Y \to X) = I(X_{t+1}; Y_{t-\tau}) - I_{min}(X_{t+1}; Y_{t-\tau}, X_{t-\tau}) \tag{4}$$

$$SDTE_\tau(Y \to X) = I(X_{t+1}; Y_{t-\tau}, X_{t-\tau}) - I_{max}(X_{t+1}; Y_{t-\tau}, X_{t-\tau}) \tag{5}$$

$$TE_\tau(Y \to X) = SITE_\tau(Y \to X) + SDTE_{\tau_x}(Y \to X) \tag{6}$$

where I_{min} is defined as:

$$I_{min}(X_{t+1}; Y_{t-\tau}, X_{t-\tau}) = \sum_{x_{t+1}} p(x_{t+1}) \min_{R \in \{Y_{t-\tau}, X_{t-\tau}\}} I(X_{t+1} = x_{t+1}; R) \tag{7}$$

and I_{max} is defined the same way except substituting min with max.

In order to remove the bias due to the statistical properties of the time series, and in order to make the information transfers between different signals comparable, we subtract the shuffled information transfer and normalize it to the range $[0, 1]$ according to [11]. The shuffled information transfer is calculated by first scrambling the data of the time series Y so that the time-dependency is lost but the statistical properties remain. The normalized transfer entropy is then expressed as:

$$TE_\tau(Y \to X) = \frac{TE_\tau(Y \to X) - TE_\tau^{shuffled}(Y \to X)}{H(X_{t+1}|X_{t-\tau})} \tag{8}$$

2.2. *Experimental Setup*

The experimental setup was identical to our previous work [32]. We recapitulate it here for the reader's convenience.

2.2.1. *Robotic Platform and Control Signals*

The robot used (*Fig.* 1 (a)) had four identical legs driven by position-controlled servomotors in the hips. It had passive compliant joints at the knees. Upper and lower limb were connected with springs. A special material (adhesive skin used for

ski touring from Colltex), which has asymmetrical friction properties, was added onto the robot's feet. This allowed the robot to get a good grip during leg retraction (stance), and enabling sliding during protraction (swing). The mechanical design (weight distribution, proportions, springs used, etc.) was a result of previous research (e.g. [16]).

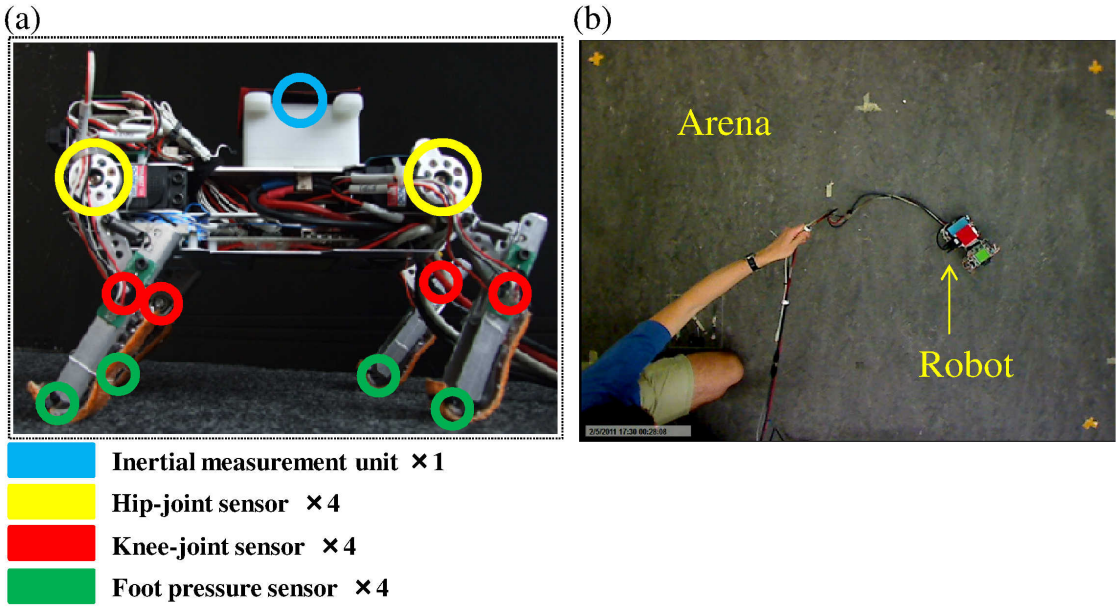


Fig. 1. **Robot experiments. (a)** The quadruped robot "Puppy" with its set of sensors (colored circles). **(b)** The arena used in the experiments (linoleum ground shown). The picture was taken from an overhead camera which was used to track the robot trajectories.

We prepared three sets of position control commands for the servomotors, resulting in three distinct gaits. The first one was the *random* gait, where the target hip joint angle for each leg was set randomly with certain smoothing constraints (to avoid too high frequencies that would exceed the motor bandwidth). The remaining two gaits were based on a simple oscillatory position control of the motors, each motor signal a sine wave. The target hip joint angle γ_i of each motor i (and hence of each leg) was determined as

$$\gamma_i(t) = \alpha_i \cdot \sin(2\pi f t + \theta_i) + \beta_i, \tag{9}$$

where the oscillation was varied by changing the amplitude α_i, offset β_i, frequency f, and phase lag θ_i parameters. Offset β_i defines the center of the oscillation. In the experiments reported here, frequency f of all legs was set to 1 Hz. By experimentation, we have prepared two parameter settings which gave rise to two turning gaits. The *bound right* gait was derived from a bounding gait; the *turn left* gait achieved the left turn by simply using a higher amplitude in the hind right leg. The motor signals of the three gaits are shown in *Fig.* 2.

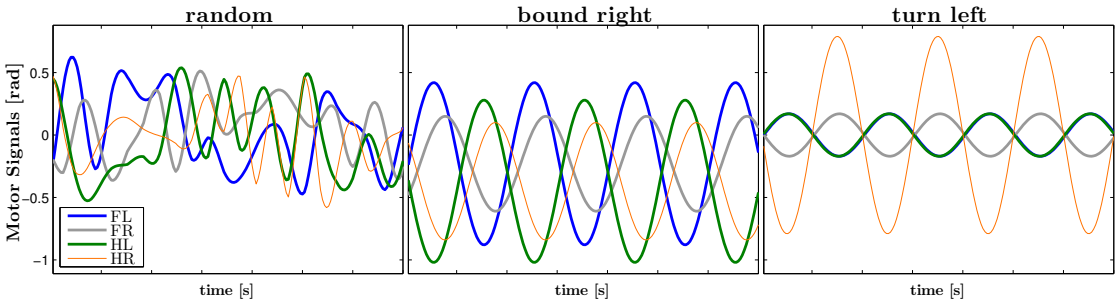


Fig. 2. **Motor time series.** The plots show 3.5 s of the motor commands as the robot runs with *random*, *bound right* and *turn left* gait respectively. The signals are shown for every leg, *FL*: front left, *FR*: front right, *HL*: hind left, *HR*: hind right.

Besides the four motor channels (denoted as M_{FL},M_{FR},M_{HL},M_{HR}), we used 18 sensory channels from the robot (see *Fig.* 1 (a)). Eight potentiometers were used to measure the joint angles, four on the active hip joints (H_{FL},H_{FR},H_{HL},H_{HR}) and four on the passive knee joints (K_{FL},K_{FR},K_{HL},K_{HR}). On the robot's feet were four pressure sensors (P_{FL},P_{FR},P_{HL},P_{HR}). Linear accelerations in three axes (A_X,A_Y,A_Z) and angular velocities around the three axes (G_X,G_Y,G_Z) were taken from an inertial measurement unit (IMU). All sensory data were sampled at 50Hz. For convenience, we refer to the hip and knee angular sensors as "hips" and "knees" and speak about "motors" when we mean the motor commands.

2.2.2. *Arena and Ground Conditions*

During the experiments, the robot was running in an arena roughly 2.5 x 2.5 m and was tethered[a] (*Fig.* 1 (b)). The turning gaits were chosen to keep the robot inside the arena. To investigate the effect of ground conditions, we used five different ground materials: *linoleum*, *foil*, *cardboard*, *styrofoam* and *rubber*. The main difference was in the friction coefficient between the ground material and robot's feet[b] In addition, the *rubber* and *cardboard* contained regular ridges.

2.2.3. *Experiments*

As we have discussed in Sec. 1, behavior is an outcome of the dynamical reciprocal coupling of the brain, body and environment. *Fig.* 3 illustrates this schematically. All the interacting components introduce some constraints on the interplay and

[a]Cables were used for data transfer and power transmission. Although they did affect the robot's dynamics, an effort has been made to minimize these effects by carrying the cables by the experimenter.

[b]We estimated static friction coefficients by putting a block covered with the same adhesive skin as on the robot's feet on inclined planes covered with the different ground materials. As the adhesive skin has asymmetrical properties, two values were obtained for each material. The low/high values were: *linoleum*: 0.31/0.40, *foil*: 0.39/0.39, *cardboard*: 0.64/1.10, *styrofoam*: 0.74/1.06, *rubber*: 0.76/0.91.

together induce some regularities or structure. Adopting the situated perspective, we will study how much can be inferred by the agent about the interaction from observing the sensorimotor flows only. To this end, we have designed experiments in which two of the interacting components are systematically varied.

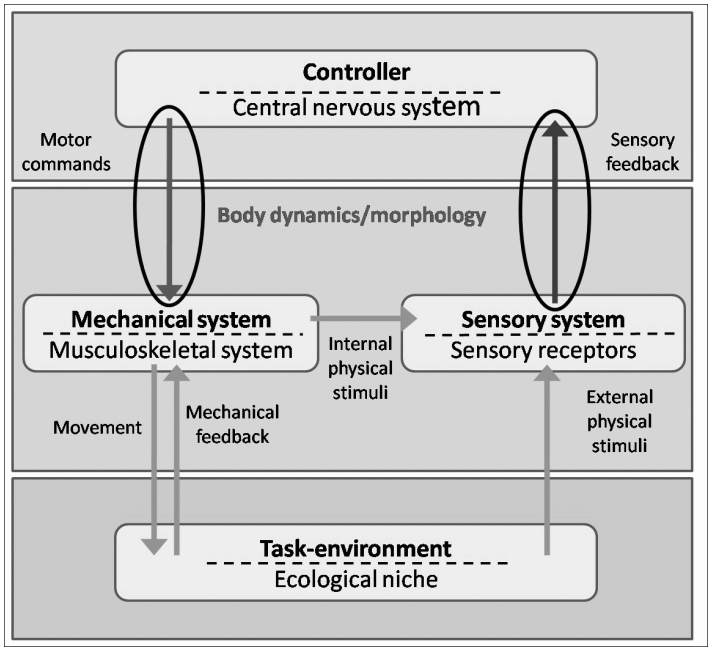


Fig. 3. **The interplay of information and physical processes.** Driven by motor commands, the mechanical system of the agent acts on the external environment (different ground substrates in our case). The action leads to rapid mechanical feedback (for example, springs in the passive knees are loaded). In parallel, external stimuli (pressure on the robot's feet or acceleration due to gravity) and internal physical stimuli (bending of joints) impinge on the sensory receptors (sensory system). The arrows marked with ellipses correspond to the information flows that are available to the agent's "brain" for inspection; these are the subject of our analysis. Figure and text adapted from [29].

Experiment 1: *Varying the controller.*

We have varied the control signals sent to the robot's motors, which give rise to disctinct gait patterns. A set of random signals (*random*) and two coordinated motor controllers (*bound right* and *turn left*) were prepared. Keeping the body and environment constant (*linoleum* ground was used), we investigated how the information structure changes with the different controllers.

Experiment 2: *Varying the environment.*

By fixing the controller to the *bound right* gait, we investigated how the ground conditions affect the informational structure experienced by the robot. For the ground conditions, we used *foil*, *linoleum*, *cardboard*, *styrofoam* and *rubber*.

The body was not varied in our experiments. However, as the other main actors - the control signals and the environment - were systematically manipulated, it was to some extent possible to investigate its effect by uncovering the invariant (always present) structure in the sensorimotor space.

2.3. *Data Analysis*

The trial durations of the experiments were between 60 and 130 seconds. To have an equal number of samples for the calculations of the information transfer, we divided longer trials into subtrials of 58s length (2900 samples). Additionally, we discarded the first 2 seconds (100 samples) of each trial, in order to exclude the data of the transition from sitting to running. This way we obtained between 2 and 5 subtrials per condition. The marginal and joint probability distributions that were needed to calculate the information flows were estimated using histograms. After normalizing the time series to a standard normal distribution ($X, Y \sim \mathcal{N}(0, 1)$), the state space was divided into 20 equally spaced bins ranging from [-4, 4] and the frequency of each state was counted. We tried different bin numbers (5 to 64) and ranges and observed no qualitative difference in the resulting information transfer. The information transfer was then averaged over all trials in each condition. We calculated them for time lags $\tau = [0, 1]$ seconds (1 second was the period of locomotion of the robot) and selected the maximum across $\tau = \text{argmax}_\tau \left[TE_\tau(Y \to X)\right]$. The shuffled information transfer used for the normalization was calculated by scrambling the time series of Y 100 times, then calculating the information transfer for all 100 scrambled time series and taking the mean of that.

3. Results

In this section, we start by briefly looking at the behavior of the quadruped robot in the arena from an observer perspective. Then we will analyze the information structure that can be extracted from the time series. We will see how the behavior is reflected in the information structure and how it can contribute to the robot's perception and development. We also would like to draw the reader's attention to a video from the experiments that will provide a clearer picture of the experimental conditions in which the robot interacts with its environment: `https://files.ifi.uzh.ch/ailab/people/hoffmann/videos/ACS2012/SchmidtEtal_ACS_2012_accompVideo.mpg` or `.wmv`.

3.1. *Behavior*

Fig. 4 (a) compares example trajectories of the robot for the three different gaits on the *linoleum* ground. We can clearly see that the robot's behavior was different in each gait. Interestingly, we found that for the *random* gait the robot moved forward and had a tendency to turn clockwise (light gray line). Since the motor signal was random, we attribute this pattern to the asymmetry in the morphology

of the robot. The forward motion can be attributed e.g. to the mass distribution, the leg shape, and the asymmetric friction properties of the adhesive skin on the robot's feet. The turning effect could be explained by the IMU attachment with the cable pointing to the right. In the *turn left* gait, the trajectories of the robot showed counterclockwise circles (dark gray line). From the pronounced "zig-zag" shape - as seen by the overhead camera - we can also observe that the robot's rolling motion was substantial and larger than the forward motion of the robot in each locomotion period. In the *bound right* gait, the robot turned clockwise (black line). The diameter of the trajectory seemed to be modulated by the ground type (*Fig.* 4 (b)). In particular, if the friction between ground and feet was larger, the diameter of the circle was smaller. In summary, the robot showed a characteristic behavior for each gait condition (controller) and its behavior is strongly affected by the ground type (environment).

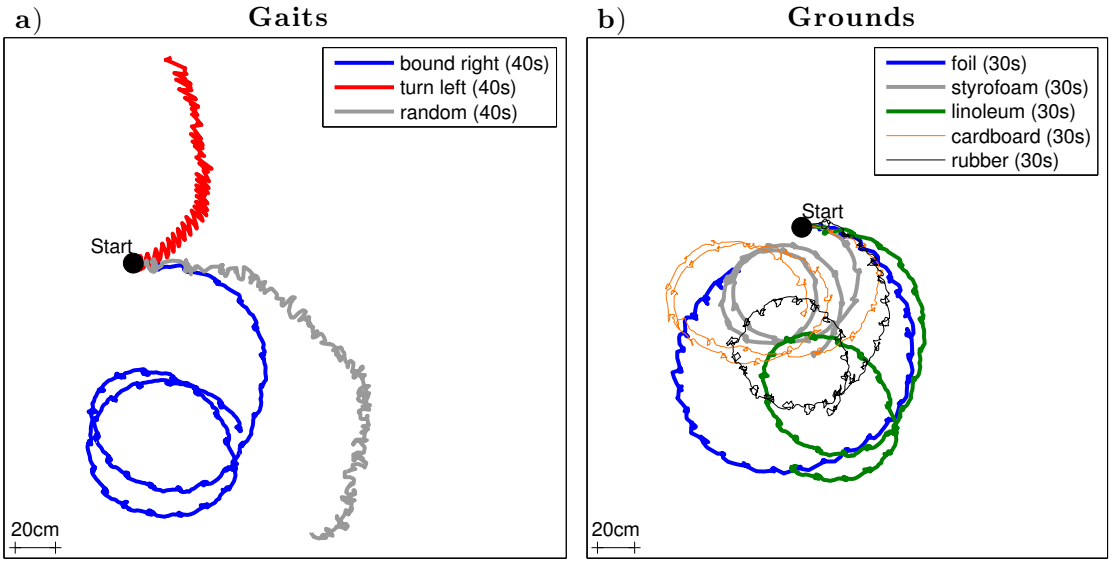


Fig. 4. **Robot trajectories.** Typical trajectories of the robot's center of mass in the arena, as viewed from above. **(a)** The trajectories on the *linoleum* ground when running with *bound right* (black), *turn left* (dark gray) and *random* (light gray) gait. **(b)** The trials from different grounds with the *bound right* gait. The trajectories also reveal how the turning radius and the distance traveled (the speed of the robot) was dependent on the ground condition.

3.2. *Experiment 1: Influence of the Controller on the Information Structure*

3.2.1. *The Random Controller and Body Schema Synthesis*

We start by analyzing the *random* controller, in which the motor commands were set randomly and independently so that there was no correlation among them. If we let the robot run long enough, in the limit we will encounter all possible

combinations of motor commands of the four legs. The *random* controller can then be seen as marginalizing out the controller part of the controller-body-environment system. Hence, information structure obtained with this gait can be considered to be induced by the interaction between the body and the environment of the robot only. *Fig.* 5 shows the transfer entropy among all the variables in the *linoleum* ground condition. A cell of the matrix (a) indicates the information transfer from the signal in the column to the signal in the row.

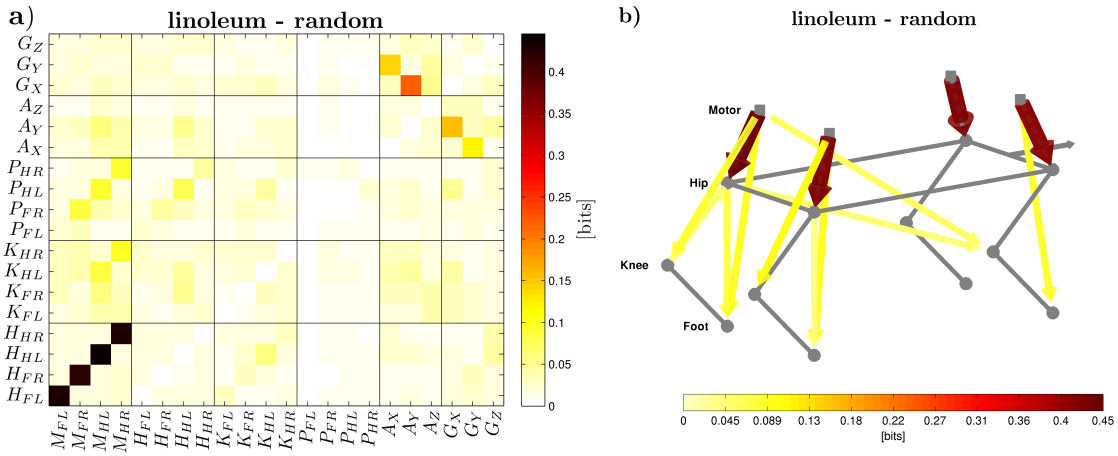


Fig. 5. **Transfer entropy TE in the *random* gait on *linoleum*. (a)** Every cell of the matrix corresponds to the information transfer from the signal on the column position to the signal on the row position. **(b)** A schematic of the Puppy robot (dashed lines) with overlaid arrows depicting the TE between the individual components. For readability, only the 15 highest values are shown and the accelerometers and gyroscopes were excluded from this visualization. The strength of the information transfer is encoded as thickness and color of the arrows.

The strongest information transfer occurs from the motor signals to their respective hip joint angles ($M_{FL} \rightarrow H_{FL}$, $M_{FR} \rightarrow H_{FR}$, $M_{HL} \rightarrow H_{HL}$, $M_{HR} \rightarrow H_{HR}$). The motors directly drive the respective hip joints and, despite some delay and noise, the hip joints always follow the motor commands, which induces a strong informational relation.

The motors further show a smaller influence on the knee angles (especially at the hind legs K_{HL} and K_{HR}) and on the feet pressure sensors, all on the respective leg where the motor is mounted. Finally, also the hip joints have some weak influence on the pressure sensors of the respective leg. The schematic of the Puppy robot in *Fig.* 5 (b) shows the same information flows as arrows with thickness and color depicting the strength of information transfer between the components (sensors from the IMU are not shown in this schematic). It can be seen that the information is mainly propagated within each leg, with stronger flows in the hind legs.

Other interesting relations revealed by the transfer entropy are between A_Y and G_X, and between A_X and G_Y. These reflect the robot's pitching and rolling movements, respectively, which are prominent motions in the quadruped robot.

When the robot rolls to one side, the gyroscope measures angular velocity around the X-axis (G_X), while the acceleration due to gravity partly projects into the Y-component, appearing in A_Y. Similarly the pitching movement affects the sensors A_X and G_Y.

Concluding, while the overall information in the gait induced by the *random* controller is quite low, the few relations that stick out reflect many things we know about the robot's physical structure and its behavior. In particular, the information flows between sensors and motors of the same leg are prominent, and the rolling and pitching movements induce flows between accelerometers and gyroscopes.

We propose that the contingencies derived from this gait constitute a rudimentary body representation of the robot. Of course this body schema is only valid in the environment the robot has experienced during the trials (*linoleum* in this case, but could be extended to all available ground conditions). We want to emphasize that contrary to the work in robotics dealing with self-recognition or self-calibration that we have reviewed in Sec. 1, the agent can arrive at this model with minimal assumptions or prior knowledge.

3.2.2. *Coordinated Motor Commands*

In the following, we will see how the information structure changes if we introduce controllers with coordinated, synchronized motor commands, which give rise to the *bound right* and *turn left* gait. The motor commands in these gaits are periodic oscillatory signals of the same frequency, but of different amplitudes, offsets and phase. Consequently, the robot exhibits periodic behavior and periodic-like signals are induced in the sensory channels.

From *Fig.* 6 (a, c) we see that the overall amount of information transfer is much higher than with the *random* controller. Furthermore, the information transfer no longer occurs only among the variables within one leg, but also among variables belonging to different legs. In the *bound right* gait (a, b), all motors (M) transfer much information to all hip joints (H). The knee joints receive as much information from the motors as the hip joints. In the *turn left* gait (c, d), on the other hand, the strongest influence of the motors is on the hind right hip (H_{HR}), followed by the hind knees. The pressure sensor P_{HR} receives more information than the other pressure sensors and the flows among different hip joints are very low except for the flows from the H_{HR} to the others. The special role of the hind right leg in the flows reflects the fact that the M_{HR} motor has a much higher amplitude and is a key contributor to the robot's locomotion (cf. video).

The information flows from the motors to the inertial sensors can be also related to the behavior displayed by the two gaits. The high flows from the motors to G_Y in the *bound right* gait relate to the pitching movement. The trajectories as observed by the overhead camera (*Fig.* 4 (a)) show a pronounced sidewards "zig-zag" movement during the *turn left* gait. This corresponds to the roll motion which is reflected in

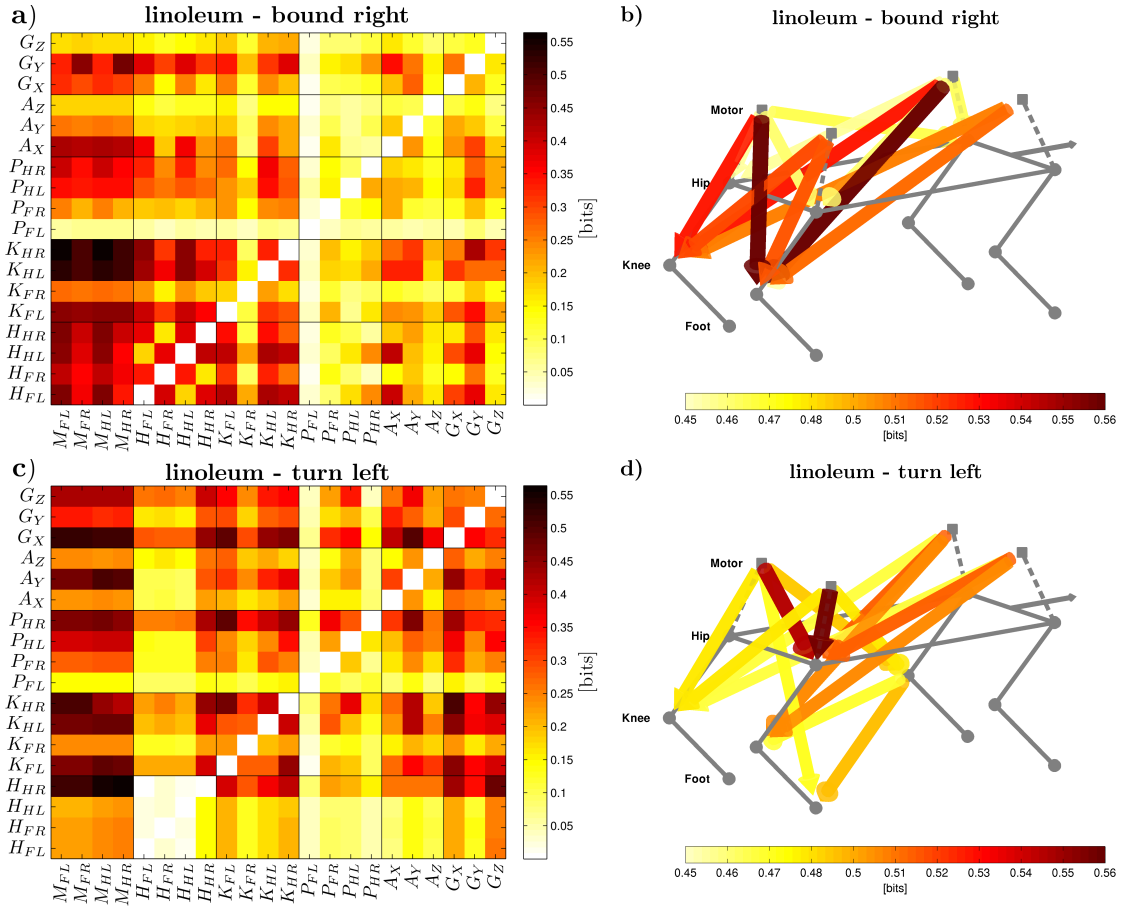


Fig. 6. **Transfer entropy in the coordinated gaits on *linoleum*.** The matrices and Puppy schematics show the transfer entropy in the *bound right* gait **(a,b)** and the *turn left* gait **(c,d)** in the same way as described in *Fig.* 5. Please note the different scale of the matrices and the Puppy schematics.

the high flows from the motors to A_Y and G_X.

All these examples show that the measured information flows strongly reflect aspects of the robot's behavior and the physical properties of its body. Furthermore, they show that periodic behavior induces specific informational structure through synchronization.

3.2.3. *Controllability*

Can maps of sensorimotor flow be utilized for control purposes, i.e. to achieve desired states or goals by the robot? As stated in Sec. 2.1, the transfer entropy from a controller to a variable expresses the controllability of this variable. Moreover, the decomposition into state-independent and state-dependent transfer entropy ($SITE$ and $SDTE$) allows to distinguish the open-loop and closed-loop controllability of the variable. This means the agent can infer the controllability of its sensory channels by its motors by looking at the flows from its actuators to its sensors. *Fig.* 7

shows this decomposed directed information transfer for the three sets of motor commands used in the Puppy robot. Part a) - from the random controller - hints on the controllability of the platform in general. We see that the hip joints can be controlled by the motors in the respective leg and there is indication that this can be done in an open-loop fashion, since the $SITE$ component is stronger [c]. The flows to the knee and pressure sensors, in particular in the hind legs, also hint on their possible controllability in an open-loop fashion.

Fig. 7 b) and c) depict the situation of the coordinated gaits. The information flows indicate higher open-loop controllability of the hips and the pressure sensors (stronger $SITE$ part). Although we saw in the previous section that in the coordinated gaits the knees receive more total information from the motors than the hips, the decomposition shows that this mainly comes from the $SDTE$ – their closed-loop controllability. So in order to control the knees, the feedback about their current state may be needed.

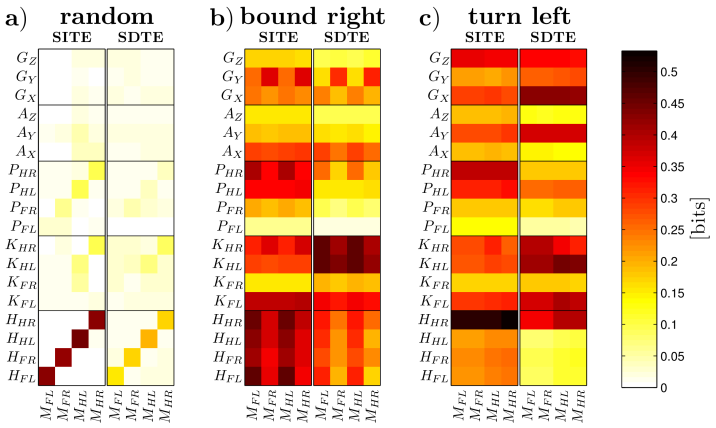


Fig. 7. **Decomposition of the transfer entropy.** The matrices show the decomposition of the information flows observed from the three controllers into $SITE$ and $SDTE$. Only the flows from the motors to the sensors of the robot are shown.

What would be the first goal-oriented behaviors that would be meaningful in the current situation? Let us imagine that the robot "wants" to accelerate forward or to turn. That is, a desired sensory state would be a high value of A_X or G_X respectively. From the information flows, we see that A_X is more affected by the motors in the *bound right* gait, whereas for G_X, it is the *turn left* gait that shows a stronger flow. The robot could thus choose a gait that would make the desired control action easier. Then, to obtain a simple controller, we would be interested in the "inverse" mapping - from the sensory variable to the motor signal - that would

[c] In reality, every hip is controlled with a closed-loop controller of the servomotor. However, this is hidden from the robot and hence, from a situated perspective, it is plausible to assume that they can be controlled in open-loop.

give us appropriate motor commands. This mapping could then be worked out as a functional relationships using regression, for example.

3.3. *Experiment 2: Influence of the Environment on the Information Structure*

In experiment 2 we let Puppy run with the *bound right* gait on five different grounds to investigate how the interaction with different environments changes the information structure.

3.3.1. *Ground Discrimination*

Fig. 8 shows the standard deviation of the information flows across the five ground conditions (after averaging the trials within the ground conditions). It reveals which flows are sensitive to changes of the ground and which remain constant. The matrix (a) shows that especially among motors and hip joints the relations are very strong and invariant to ground changes. On the contrary, the information that the hind left pressure sensor P_{HL} receives from many of the other channels, is very dependent on the ground condition, which can be seen from the arrows to P_{HL} in (b). We can again see that the flows reflect some properties of the robot's body: the hips follow the strong motors and are largely unperturbed by variations of the environment, while the pressure sensor measures directly the ground contact and can sense the differences (especially the hind left one, which takes the highest load during forward-rightward pushing in this gait).

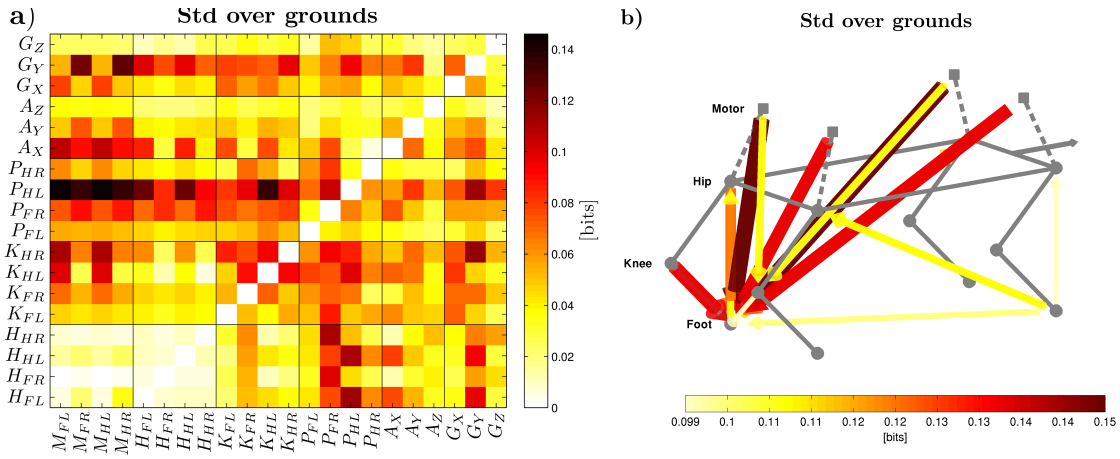


Fig. 8. **Information flow on different grounds.** The matrix **(a)** and Puppy schematic **(b)** show the standard deviation of the transfer entropy across the five ground conditions while running with the *bound right* controller. The standard deviation is calculated after averaging the trials within each condition.

We extended the analysis of variation induced by the ground conditions by a principal component analysis on the flows in all trials. The matrix in *Fig.* 9 (left)

shows the resulting 1^{st} principal component of the transfer entropy. It shows that a lot of variation comes from the influence of the motor commands on the sensory channels. Especially the left knees (K_{FL}, K_{HL}), the pressure sensor P_{HL}, the accelerometer A_X and the gyroscope G_Y receive different information in different trials.

Plotting the information flow of all trials of the five ground conditions in the space spanned by the first two principal components (*Fig.* 9 (right)), confirms that the highest variance directions are indeed separating the ground conditions very well, and that trials on the same ground are clustered. This shows the robot's capability to distinguish the environmental conditions by observing the changes in certain information flows.

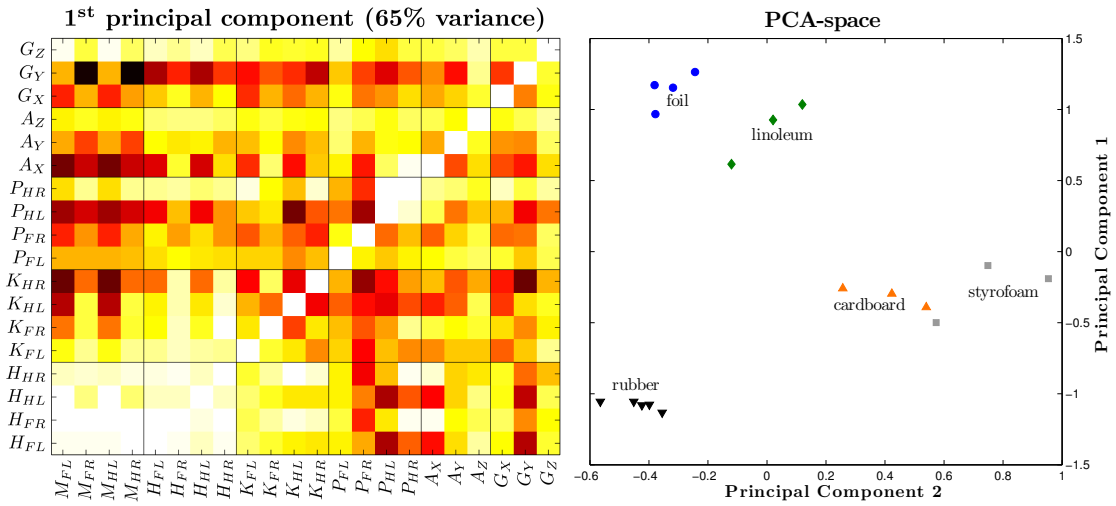


Fig. 9. **PCA of the information flows. (left)** The matrix shows the 1^{st} principal component of the information flows in all trials of the five grounds conditions while running with the *bound right* controller. **(right)** Information flow of all the trials projected onto the the first two principal components and labeled according to their ground condition.

3.3.2. *Stability and Friction*

Fig. 10 (b) shows the mean information transfer over all sensor and motor pairs depending on the friction coefficient estimate between each ground material and Puppy's feet. The amount of information transfer is negatively correlated with the friction coefficients ($r = -0.88$).

The dashed line in the figure shows a stability measure of the robot's locomotion. It measures the variation of all sensory channels from one period of locomotion to the next (perfectly periodic signals mean perfectly stable locomotion), and is highly correlated with the overall information transfer ($r = 0.995$). This also matches with an outside inspection of the robot's smoothness or comfort of locomotion, which is

extremely smooth on the *foil* ground and becomes very difficult on higher-friction grounds, especially on *rubber* (cf. video). Therefore, the mean information transfer could serve as a possible reward or cost function that the robot could try to optimize - choosing the gait that has the highest score on a given terrain, for instance. Further explorations in this direction are necessary and could draw from existing work in this area [5, 27, 31, 36, 17].

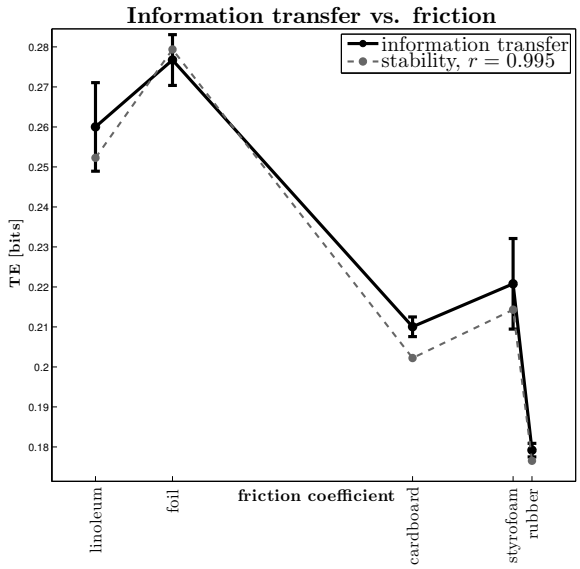


Fig. 10. **Information transfer vs. friction.** The mean information flow across all variable pairs is plotted against the static friction coefficient estimate in each ground condition (black solid line with error bars). In addition, the stability (dashed gray line) is depicted. The stability is calculated from the standard deviation of all sensor signals across locomotion periods (a perfectly stable behavior would be perfectly periodic and have zero variation, whereas an unstable behavior would show some variation and thus have a negative value in this stability measure).

3.4. *Sensorimotor Contingencies*

3.4.1. *Proprioceptive and Exteroceptive Sensors*

Sensors that an agent possesses are often classified into proprioceptive and exteroceptive. For robots, Siegwart et al. [35] define *proprioceptive sensors* as those that measure values internal to the robot (e.g. battery voltage, joint angle sensors) and *exteroceptive sensors* as those those that acquire information from the robot's environment (e.g. distance sensors, cameras). However, these classifications rely on an *a priori* knowledge about what is internal to the agent and what is external environment. We will assume that this is not known to our robot, the agent is only confronted with the signals reaching its "brain".

Philipona et al. [30] define the agent's body as part of the world over which it has complete control. Consequently, proprioceptors are defined as input channels

with high controllability. We adopt the notion of controllability from [42] as being quantified by the information transfer from a controller (motor commands in our case) to a variable (sensors in our case). In this sense, "proprioceptiveness" is not an all-or-nothing classification of a sensor, but rather a continuous property. *Fig.* 11 (left) visualizes this for Puppy's sensors. It shows the information flow from the four motors (M_{FL}, M_{FR}, M_{HL}, and M_{HR}) to each sensory channel. We can immediately spot that the hip angular sensors stand out. They receive very high information flows from the respective motor signal of the same leg. Thus, the agent could attribute the proprioceptive property to the hip potentiometers. While the other sensors do not reach as high values, some degree of "proprioceptivness" can still be observed. We can see that the knee and pressure sensors also receive significant information flows from their respective motor signals on the same leg.

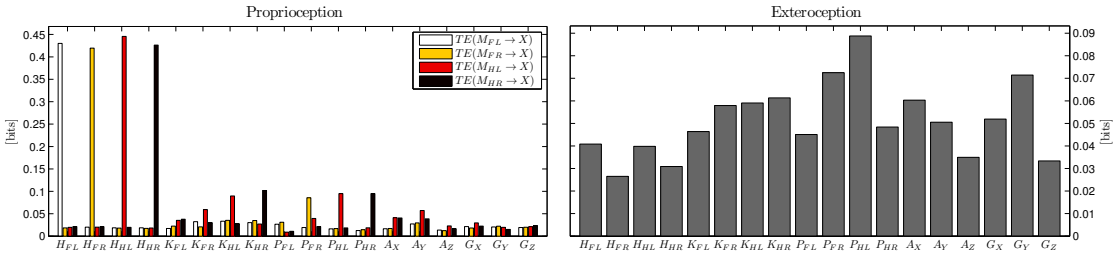


Fig. 11. **Proprioceptive and exteroceptive sensors. (left)** Proprioception is defined as the information flows from motor signals to each sensory channel under the *random* gait on the *linoleum* ground. The four bars in each sensory channel represent the flows from the M_{FL}, M_{FR}, M_{HL}, and M_{HR} motor signal respectively. **(right)** Exteroception is defined as an aggregate measure of how the information flows to and from each sensor vary when the ground is varied (standard deviation across five grounds with the *bound right* gait).

Exteroceptors can be defined as the channels that are sensitive to changes in the environment. In Sec. 3.3 and *Fig.* 8 we have shown that the information flow between each motor-sensor or sensor-sensor pair varies when the ground changes. By averaging over this standard deviation of all incoming and outgoing flows of a channel (row and column involving this channel), we can estimate the overall level of proprioception of each channel individually. *Fig.* 11 (right) shows the "exteroceptiveness" for each channel and it shows that this is again not an all-or-nothing property, but a graded distinction of the channels.

Compared to classical sensor classification, where angular and inertial sensors are classified as proprioceptors and tactile (pressure) sensors as exteroceptors, our interpretation derived from the information structure provides a very different picture, as can be inspected in *Fig.* 12 (left), where the two sensor characteristics are plotted against each other. Whereas the hip angular sensors are clearly identified as proprioceptors, their "colleagues" in the knee joints show both properties to a similar extent. In the context of our robot, we find this plausible. While the hips are

directly driven by motors, the knee joints are passive and are also highly dependent on the interaction with the ground. Similarly, the inertial sensors seem to be more responsive to environmental changes than to the individual motor signals.

We want to argue that the sensor classification as we have just demonstrated reflects much more the reality as experienced by the agent than the classical textbook classification would.

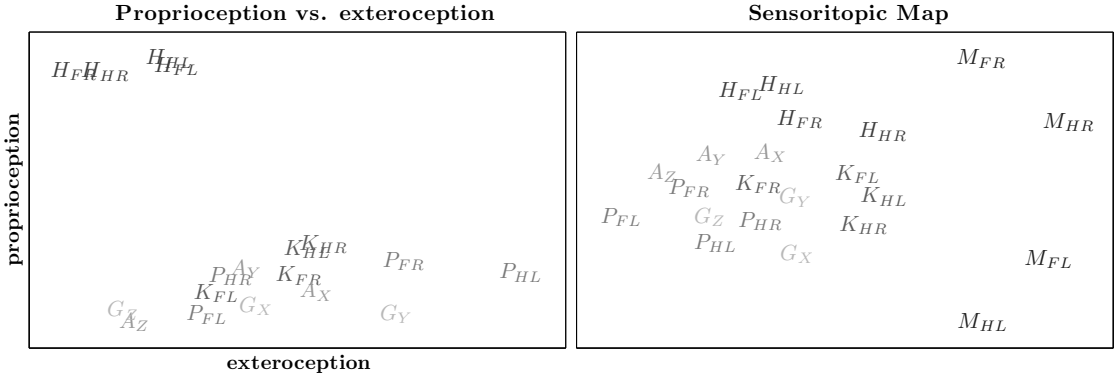


Fig. 12. **Sensor spaces. (left)** Proprioception vs. exteroception. The values from *Fig.* 11 are plotted against each other. **(right)** A Sensoritopic map. Projection of the sensors and motors into 2D space using multidimensional scaling based on a information flow-based similarity measure.

3.4.2. *Learning about Sensory Modalities*

According to O'Regan and Noe [26], it is the "structure of the rules governing the sensory changes produced by various motor actions" what differentiates modalities. The directed information flows that we have quantified provide the basis for such a structure. The agent could assign two channels that are similar to a common modality. Similar in terms of information flows means that they send and receive same amounts of information to/from the same channels[d]. In *Fig.* 12 (right) we show how such an information flow-based similarity measure leads to a map of the agent's sensor space, a sensoritopic map, by projecting the channels onto a 2D plane where their distance reflects their similarity[e]. The resulting map shows a reasonable clustering of channels belonging to same modalities. In particular, the motors are located on the far right, the hip joint angles come together at the bottom, the knee joint angles central and the pressure sensors on the left. The inertial sensors are scattered between the knees and pressure sensors towards the top. The lack of topological relationships (sensors of the same leg do not come together) comes probably from the fact that in our platform, there are few separated physical

[d]This distance metric used here is the Euclidean distance between the rows and columns of two sensors in the information transfer matrices.
[e]This was achieved by multidimensional scaling, similar to what has been done in [25].

relationships. As the robot runs, through the interaction with the environment, the influence of one leg gets propagated to all the other legs.

3.4.3. *Predictive Capacity*

Transfer entropy measures how knowing the state of one channel helps in predicting the state-transition of another channel. Thus averaging all the values in a column of a transfer entropy matrix will give us an aggregate "predictive capacity" of each channel. The predictive capacity can serve as an indicator of the channel's quality or utility for the agent. The result of this analysis is depicted in *Fig.* 13. Not surprisingly, the motor signals have the highest score. They are controlling the system and should thus be most effective in predicting the sensors' future states. Among the sensory channels, some hip and knee angle sensors have high values. These are good candidates to focus attention on. Conversely, sensors with low scores (e.g. P_{FL}) could receive less attention or be marked for replacement.

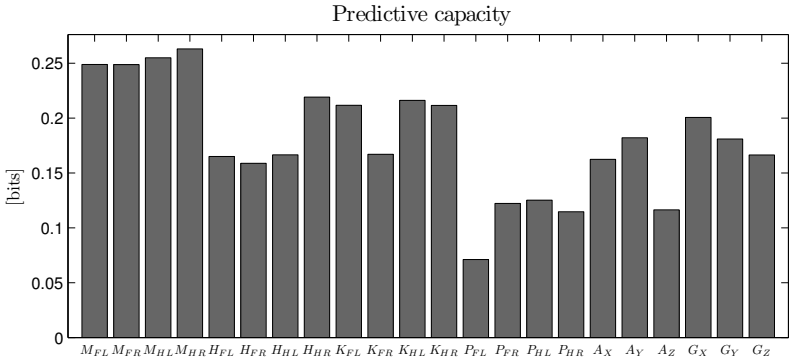


Fig. 13. **Predictive capacity.** The mean information transfer from each sensor or motor to all other channels (mean across all gaits and the ground conditions *foil*, *linoleum* and *styrofoam*).

4. Discussion

The measure we used to analyze the information flows in time series was transfer entropy. In addition, in the analysis of the information transfer from the motors to the sensors, we have employed its decomposition into state-independent (SITE) and state-dependent transfer entropy (SDTE), motivated by its relationship to open-loop and closed-loop controllability [42]. Transfer entropy fulfilled our criteria as a method capable of extracting directed nonlinear relationhips. Other information theoretic methods could possibly be applied, however, a quantitative comparison of different methods was not the purpose of the current study (see [9] for a study along these lines). Nevertheless, all methods that rely on observing time series only have difficulties separating real causal effects from spurious correlations. There,

interventional methods [28, 1] could be used to refine the relationships that we have extracted.

We have used a real, dynamic, nonlinear platform equipped with 18 sensors encompassing multiple sensory modalities. We want do discuss a number of points regarding this choice. First, in our view, this platform bears a fair level of ecological validity and would satisfy what Ziemke [44] has called "organismoid embodiment": an organism-like bodily form with sensorimotor capacities akin to living bodies (though our dimensionality is still much lower compared to sensorimotor spaces in biology). This contrasts with the studies on artificial agents in which highly simplified abstract worlds are often used. Second, the nature of legged locomotion - a periodic behavior composed of alternating phases of leg touchdown and lift-off - poses specific challenges. Special care needs to be taken when applying information transfer analysis to periodic signals. The contacts with the ground, on the other hand, introduce sharp discontinuities in the dynamics. This contrasts with robotic case studies, where the environment is sampled by a smoothly moving camera. Third, distal or "visual" sensors are completely absent in our case. Hence, our robot cannot see itself and thus can obtain almost no information about its state while being static. Active generation of information is thus indispensable and the sensorimotor flows that we analyze uncover complex implicit dynamic relationships in a running legged robot, rather than straightforward geometrical transformations.

Let us look at our case study from an engineering perspective as well. The relationship between motor and sensory signals that we have analyzed would fit into the scope of system identification methods (e.g. [18]), possibly giving rise to a model of our robot (the plant). This could be of a grey-box, where knowledge about the system would enter the model, or black-box kind - corresponding to our situation, where the agent has little prior knowledge regarding its body, environment and nature of actuators and sensors. In fact, the input signals that we have used in our scenario - periodic and random motor signals - correspond to possible ways of exciting a system in open loop from system identification [18, 12]. The random motor signal has the advantage of being "rich" enough - containing many frequencies. In addition, it does not contain any information structure in it, which proved very useful in our situation - we have found that the information theoretic analysis is very sensitive to structure from a periodic motor signal being "imposed" on the sensory signals. Can our study inform the system identification community? We propose that an information-theoretic analysis of the kind we have performed could act as a first step that would reduce the dimensionality of the problem and point to the important relationships which can be later modeled in detail (using transfer functions, for instance).

5. Conclusion and Future Work

In this paper, we have analyzed the sensorimotor flows in a running quadruped robot using transfer entropy and studied the impact of different environmental conditions

as well as motor signals on the information flows. Then, we have adopted a situated perspective (looking at the world through the "eyes", i.e. sensors, of the autonomous agent) and proposed a number of ways in which the agent could use the tools of information theory to discover the regularities in the sensorimotor space that its interaction with the environment induces and use these to bootstrap its perception and cognition.

We see a lot of potential for future work in both the analytical and the application part of our case study. In the analysis, first, we have looked at information flows between pairs of variables (motor or sensor) only. The method could be extended to multivariate information transfer as suggested in [42]. Second, in the analysis of pairs of time series, we have collapsed the time dimension by selecting the time lag at which the information transfer was maximum. However, the exact timing of the information transfer is important for control purposes as well as for perception and cognition, as demostrated by Williams and Beer [42] in a simple evolved agent. Third, we have applied only simple tools to analyze and visualize the sensorimotor structure. However, graph theory would offer further machinery that would be relevant. One could generate subgraphs based on connected components - these would correspond to local relationships, such as the motors and sensors in one leg of the robot.

We have outlined a number of directions in which understanding the structure of the sensorimotor space could bring behavioral advantage to the robot. We feel that it will be fruitful to elaborate these scenarios more concretely and put them to test. First, we have touched on the controllability in Sec. 3.2.3, where we have identified motors that could be used to control some target sensory variables. In order to acquire a controller, the knowledge that a control relationship exists, needs to be converted to a functional relationship. Edgington et al. [6] propose a method in this direction based on turning joint probability distributions into regression functions. In this way, a simple open-loop controller could be directly obtained. Second, we have shown how different environments induce changes in the flows in the sensorimotor space. In a different study on the same platform [15], we have employed sensory features to discriminate different grounds. In the future, it would be interesting to compare these results with features that use information flows instead (we show first results in Sec. 3.3.1). These features may prove to be more robust as they better reflect the overall dynamics of the robot interacting with the ground. Third, we propose that the agent can exploit the knowledge about the structure of the sensorimotor space to economically allocate its computational resources. The predictive capacity measure we have introduced in Section 3.4.3 is a rough approximation of a sensor's utility or quality that can provide a useful bias to guide the agent's attention. Furthermore, having such a measure of sensor quality can be exploited further if the agent has the possibility to change the morphology of its sensors online. If the physical placement of the sensor can be adjusted, then the agent can optimize these in order to get the most information out of each sensor. If sensor values are being discretized, then the resolution can be adapted - "good"

sensors can be sampled with more bins, for instance. Finally, if the robot detects very low information flows in one of the channels, such as the front left foot pressure sensor (P_{FL} in *Fig.* 13) in our case, this may indicate a failure. Depending on its capabilities, the robot could either try to repair the sensor or it could signal its failure.

Acknowledgments

This work was supported by the EU project Extending Sensorimotor Contingencies to Cognition (eSMCs), IST-270212.

We would like to thank Michal Reinstein who participated in running the experiments and Hugo Marques for reviewing an earlier version of the manuscript.

References

[1] Ay, N. and Polani, D., Information flows in causal networks, *Advances in Complex Systems* **11 (1)** (2008) 17–41.

[2] Cover, T. and Thomas, J. A., *Elements of information theory* (New York: Wiley, 1991).

[3] De Preester, H. and Knockaert, K., *Body Image and Body Schema interdisciplinary perspectives on the body* (John Benjamins, 2005).

[4] de Vignemont, F., Body schema and body image - pros and cons, *Neuropsychologia* **48(3)** (2010) 669–680.

[5] Der, R., Steinmetz, U., and Pasemann, F., *Concurrent Systems Engineering Series Vol. 55: Computational Intelligence for Modelling, Control, and Automation*, chapter Homeokinesis: A new principle to back up evolution with learning (IOS Press, 1999).

[6] Edgington, M., Kassahun, Y., and Kirchner, F., Using joint probability densities for simultaneous learning of forward and inverse models, in *IEEE IROS International Workshop on Evolutionary and Reinforcement Learning for Autonomous Robot Systems* (2009).

[7] Fitzpatrick, P. and Metta, G., Toward manipulation-driven vision, in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2002).

[8] Friston, K., Functional and effective connectivity in neuroimaging: a syntesis, *Human Brain Mapping* **2** (1994) 56–78.

[9] Garofalo, M., Nieus, T., Massobrio, P., and Martinoia, S., Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks, *PLoS ONE* **4(8)** (2009) e6482.

[10] Gold, K. and Scassellati, B., Using probabilistic reasoning over time to self-recognize, *Robotics and Autonomous Systems* **57(4)** (2009) 384–392.

[11] Gourvitch, B. and Eggermont, J. J., Evaluating information transfer between auditory cortical neurons, *Journal of Neurophysiology* **97(3)** (2007) 2533–2543.

[12] Haber, R. and Keviczky, L., *Nonlinear system identification - input-output modeling approach*, Vol. 1: Nonlinear system parameter identification (Kluwer Academic Publishers, 1999).

[13] Hoffmann, M., Marques, H., Hernandez Arieta, A., Sumioka, H., Lungarella, M., and Pfeifer, R., Body schema in robotics: a review, *IEEE Trans. Auton. Mental Develop.* **2 (4)** (2010) 304–324.

[14] Hoffmann, M. and Pfeifer, R., *The Implications of Embodiment: Cognition and Com-*

munication, chapter The implications of embodiment for behavior and cognition: animal and robotic case studies (Exeter: Imprint Academic, 2011), pp. 31–58.

[15] Hoffmann, M., Schmidt, N., Pfeifer, R., Engel, A., and Maye, A., Using sensorimotor contingencies for terrain discrimination and adaptive walking behavior in the quadruped robot puppy, in *From animals to animats 12: Proc. Int. Conf. Simulation of Adaptive Behaviour (SAB)* (2012), [accepted].

[16] Iida, F., Gómez, G., and Pfeifer, R., Exploiting body dynamics for controlling a running quadruped robot, in *Proceedings of the 12th Int. Conf. on Advanced Robotics (ICAR05).* (Seattle, U.S.A., 2005), pp. 229–235.

[17] Jung, T., Polani, D., and Stone, P., Empowerment for continuous agent-environment systems, *Adaptive Behavior* **19** (2011) 16–39.

[18] Ljung, L., *System Identification: Theory for the User*, 2nd edn. (Prentice-Hall, Englewood Cliffs, NJ, 1999).

[19] Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G., Developmental robotics: a survey, *Connection Science* **15(4)** (2004) 151–190.

[20] Lungarella, M. and Sporns, O., Mapping information flow in sensorimotor networks, *PLoS Comput Biol* **2** (2006) 1301–12.

[21] Maravita, A., Spence, C., and Driver, J., Multisensory integration and the body schema: close to hand and within reach., *Curr Biol* **13** (2003) R531–R539.

[22] Nakajima, K., Li, T., Sumioka, H., Cianchetti, M., and Pfeifer, R., Information theoretic analysis on a soft robotic arm inspired by the octopus, in *Proc. IEEE Int. Conf. Robotics and Biomimetics (ROBIO)* (2011).

[23] Natale, L., Orabona, F., Metta, G., and Sandini, G., Sensorimotor coordination in a baby robot: learning about objects through grasping, *Progress in Brain Research* **164** (2007) 403–424.

[24] Olsson, L., Nehaniv, C., and Polani, D., Sensory channel grouping and structure from uninterpreted sensor data, in *Proc. Conf. Evolvable Hardware* (2004), pp. 153–160.

[25] Olsson, L., Nehaniv, C., and Polani, D., From unknown sensors and actuators to actions grounded in sensorimotor perceptions, *Connection Science* **18 (2)** (2006) 121–144.

[26] O'Regan, J. K. and Noe, A., A sensorimotor account of vision and visual consciousness, *Behavioral and Brain Sciences* **24** (2001) 939–1031.

[27] Oudeyer, P.-Y., Kaplan, F., and Hafner, V., Intrinsic motivation systems for autonomous mental development, *IEEE Trans. on Evol. Comp.* **11** (2007) 265–286.

[28] Pearl, J., *Causality: Models, Reasoning, and Inference* (Cambridge Univ. Press, 2000).

[29] Pfeifer, R., Lungarella, M., and Iida, F., Self-organization, embodiment, and biologically inspired robotics, *Science* **318** (2007) 1088–1093.

[30] Philipona, D., O'Regan, J. K., and Nadal, J.-P., Is there something out there? inferring space from sensorimotor dependencies, *Neural Computation* **15** (2003) 2029–2049.

[31] Prokopenko, M., Gerasimov, V., and Tanev, I., Evolving spatiotemporal coordination in a modular robotic system, in *From animals to animats: Proc. Int. Conf. Simulation of Adaptive Behaviour (SAB)* (Springer, 2006), pp. 558–569.

[32] Reinstein, M. and Hoffmann, M., Dead reckoning in a dynamic quadruped robot: inertial navigation system aided by a legged odometer, in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)* (2011).

[33] Rochat, P., Self-perception and action in infancy, *Exp Brain Res* **123** (1998) 102–109.

[34] Schreiber, T., Measuring information transfer, *Physcal Review Letters* **85** (2000) 461–464.

[35] Siegwart, R., Nourbakhsh, I., and Scaramuzza, D., *Introduction to autonomous mobile*

robots, 2nd edn. (The MIT Press, 2011).

[36] Sporns, O. and Lungarella, M., Evolving coordinated behavior by maximizing information structure, in *Artificial Life X* (2006).

[37] Sporns, O. and Tononi, G., Classes of network connectivity and dynamics, *Complexity* **7** (2002) 28–38.

[38] Sumioka, H., Yoshikawa, Y., and Asada, M., Learning of joint attention from detecting causality based on transfer entropy, *J. Robot. Mechatronics* **20** (2008) 378–85.

[39] Thelen, E. and Smith, L., *A Dynamic systems approach to the development of cognition and action* (MIT Press, 1994).

[40] Weng, J., McClelland, J., Pentland, A., and Sporns, O., Autonomous mental development by robots and animals, *Science* **291** (2001) 599–600.

[41] Williams, P. and Beer, R., Information dynamics of evolved agents, in *From animals to animats 11: Proc. Int. Conf. Simulation of Adaptive Behaviour (SAB)* (2010).

[42] Williams, P. and Beer, R., Generalized measures of information transfer, *arXiv preprint* **1102.1507** (2011).

[43] Yoshikawa, Y., Hosoda, K., and Asada, M., Does the invariance in multi-modalities represent the body scheme? - a case study with vision and proprioception -, in *Proc. of the 2nd Int. Symp. on Adaptive Motion of Animals and Machines*, Vol. SaP-II-1 (2003).

[44] Ziemke, T., What's that thing called embodiment?, in *Proc. 25th Annual Conference of the Cognitive Science Society*, ed. Kirsh, A. . (Mahwah, NJ: Lawrence Erlbaum, 2003), pp. 1134–1139.

Appendix B

Information Flow in a Quadruped Running Robot Quantified by Transfer Entropy

Reprinted from:

Schmidt, N. M., Hoffmann, M. and Nakajima, K. (2012). "Information flow in a quadruped running robot quantified by transfer entropy", In *5th International Conference on Cognitive Systems (CogSys)*, Vienna, Austria, p. 112.

Information flow in a quadruped running robot quantified by transfer entropy

Nico M. Schmidt, Matěj Hoffmann & Kohei Nakajima

Artificial Intelligence Laboratory, University of Zurich, Switzerland

Email: nico.schmidt@uzh.ch

Background

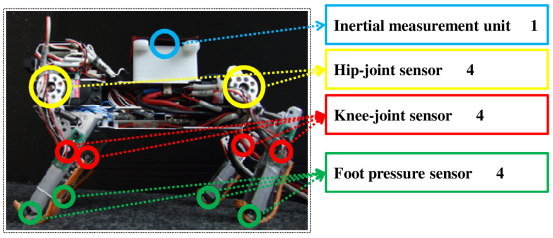
Animals and humans engage in an enormous variety of behaviors which are orchestrated through a complex interaction of physical and informational processes. The physical interaction of the bodies with the environment is intimately coupled with informational processes in the animal's brain. A crucial step toward the mastery of all these behaviors seems to be to understand the flows of information in the sensorimotor networks. In this study, we have performed a quantitative analysis in an artificial agent - a running quadruped robot with multiple sensory modalities - using tools from information theory (transfer entropy and its recently proposed decomposition). Starting from no prior knowledge, through systematic variation of control signals and environment, we show how the agent can discover the structure of its sensorimotor space. We propose that the agent could utilize this knowledge to: (i) drive learning of new behaviors; (ii) identify sensors that are sensitive to environmental changes; (iii) discover a primitive body schema.

Methods

Experiments

1. *Analysis of the information structure modulated by changing the controller.* Starting from random motor commands, we investigate how the information transfer changes when we introduce learned gaits (a bounding gait and a turning gait).
2. *Analysis of the information structure modulated by changing the environment.* We let the robot run on five different grounds (plastic foil, linoleum, styrofoam, cardboard and rubber) and investigate how the information transfer changes.

Robot Platform



Data Analysis.

Transfer entropy [1] is an information theoretic measure that quantifies the amount of information that is transferred from one signal Y to another signal X. It indicates the increase in predictability of X, if one takes the history of Y into account.

$$TE_{\tau_x,\tau_y}(\boldsymbol{Y} \to \boldsymbol{X}) = \sum_{x_{t+1}} \sum_{x_{t+\tau_x}} \sum_{y_{t+\tau_y}} p(x_{t+1}, x_{t+\tau_x}, y_{t+\tau_y}) \log \frac{p(x_{t+1}|x_{t+\tau_x}, y_{t+\tau_y})}{p(x_{t+1}|x_{t+\tau_x})},$$

References

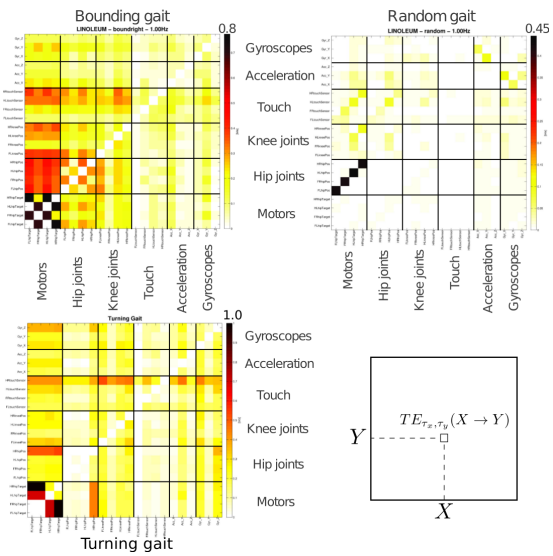
[1] T. Schreiber (2000) "Measuring information transfer". Phys Rev Lett, 85: 461–464.

[2] M. Lungarella and O. Sporns (2006) "Mapping information flow in sensori- motor networks". PLoS Comput Biol, 2: 1301–12.

This work was supported by the EU project Extending Sensorimotor Contingencies to Cognition (eSMCs), IST-270212.

Results

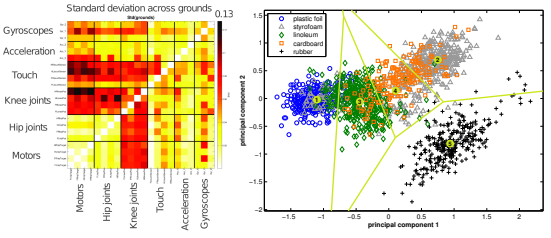
Experiment 1: Controller-induced information flow



The random gait shows only very low information transfer. The main flow is from each motor to the joint angle of the respective hip it is controlling. In the two learned gaits, the overall information flow is much higher and it has a more complex structure, including symmetric and inter-leg flows.

Question: Can the higher information flow be attributed to sensorimotor coordination (as shown in [2]) or is it mainly the structure of the control signal that is imposed on the sensorimotor data?

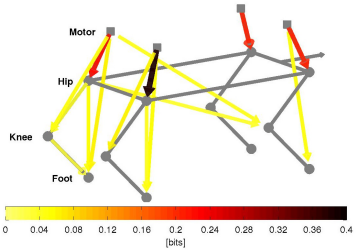
Experiment 2: Ground-specific information flow



The information transfer involving the knee joint signals or the foot pressure signals is most sensitive to changes of the ground. The ground conditions can be reliably discriminated using clustering (K-Means) of the information transfer in the subspace spanned by the first two principal components.

Conclusion: Information flow provides a new way of creating features for classification.

Information flow and body schema



The graphical representation of puppy's sensorimotor network is shown here. Overlaid is the minimal information flow that is present in all ground and gait conditions. This minimal information flow resembles the patterns that can be found in the random gait.

Conclusion: A minimal flow of information is present in all conditions and can be attributed to the physical properties of the body morphology.

Appendix C

Measuring Information Transfer in a Soft Robotic Arm

Reprinted from:

Nakajima, K., Schmidt, N. M. and Pfeifer, R. (2015). "Measuring information transfer in a soft robotic arm." *Bioinspiration & Biomimetics*, 10(3):035007. doi:10.1088/1748-3190/10/3/035007.

Measuring information transfer in a soft robotic arm

K Nakajima[1,2], N Schmidt[3], and R Pfeifer[3]

[1] The Hakubi Center for Advanced Research, Kyoto University, 606-8501 Kyoto, Japan

[2] Nonequilibrium Physics and Theoretical Neuroscience, Department of Applied Analysis and Complex Dynamical Systems, Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

[3] Artificial Intelligence Laboratory, Department of Informatics, University of Zurich, 8050 Zurich, Switzerland

E-mail: jc_mc_datsu@yahoo.co.jp

Abstract. Soft robots can exhibit diverse behaviors with simple types of actuation by partially outsourcing control to the morphological and material properties of their soft bodies, which is made possible by the tight coupling between control, body, and environment. In this paper, we present a method that will quantitatively characterize these diverse spatiotemporal dynamics of a soft body based on the information-theoretic approach. In particular, soft bodies have the ability to propagate the effect of actuation through the entire body, with a certain time delay, due to their elasticity. Our goal is to capture this delayed interaction in a quantitative manner based on a measure called *momentary information transfer*. We extend this measure to soft robotic applications and demonstrate its power using a physical soft robotic platform inspired by the octopus. Our approach is illustrated in two ways. First, we statistically characterize the delayed actuation propagation through the body as a strength of information transfer. Second, we capture this information propagation directly as local information dynamics. As a result, we show that our approach can successfully characterize the spatiotemporal dynamics of the soft robotic platform, explicitly visualizing how information transfers through the entire body with delays. Further extension scenarios of our approach are discussed for soft robotic applications in general.

1. Introduction

Recently, soft materials have been widely used to incorporate flexible elements into robots' bodies. These robots are called *soft robots* and have significant advantages over traditional articulated robots in terms of morphological deformability and interactional safety [1]. They can adapt their morphology to unstructured environments and carry and touch fragile objects without causing damage, which makes them especially applicable for rescue and interaction with humans [2, 3]. In addition, in [4], it has been demonstrated that they can generate complex behaviors with simple types of actuation by partially outsourcing control to the morphological and material properties of their soft bodies. This is enabled by the dynamic coupling between control, body, and environment, which is enhanced by the soft and flexible body [5, 6, 7].

One important difference between conventional rigid robots and soft robots is found in their body dynamics. In general, soft body dynamics exhibit diverse properties, such as high-dimensionality, nonlinearity, and a certain amount of "sluggishness" with memory, due to their elastic nature [2]. These properties usually make the robots difficult to control with methods from conventional control theory. However, in nature, some animals have soft bodies and control them in a sophisticated manner. The octopus serves as an extreme example [8]. It does not have any rigid components in its body, which can execute in virtually infinite degrees of freedom. Its motion control is far reaching in terms of conventional control framework [9, 10]. Accordingly, roboticists have been investigating how the octopus enables its motion control, searching for a novel control principle for soft robot control [3]. For example, a soft robotic arm has been proposed, inspired by the characteristic muscular structure of the octopus called *muscular hydrostat*, which enables behaviors such as reaching and grasping [11, 12, 13]. Furthermore, exploiting a material characteristic of a soft body, a robot that can crawl toward a target position in an underwater environment by coordinating the arms has been proposed in [14, 15]. A novel control scheme inspired by the organization of the octopus's nervous system has been also proposed in [16], and has been tested and implemented both in a simulator [17] and a physical platform [18, 19]. In addition, it has been recently demonstrated that the characteristic structure of the octopus arm can serve as a computational resource, which can be used as a part of a computational device [20, 21]. All of these examples are motivated to positively exploit the material characteristics of the octopus body and the diverse dynamics that the soft body generates. Based on this context, in this paper, we propose a scheme that effectively and quantitatively characterizes these soft body dynamics.

Our scheme presented in this paper is based on the information-theoretic approach. This approach is capable of analyzing nonlinear systems, and does not require precise information about the focused system beforehand, which is appropriate when the precise equations of the system are unknown (this is called a model-free approach). Due to these properties, it has been actively applied in a wide range of research fields, such as physics, neuroscience [22], economy [23], and geological physics [24, 25, 26]. In robotics,

to characterize the interaction modality of the dynamical coupling between controller, body, and environment, information-theoretic measures, such as *mutual information* (MI) and *transfer entropy* (TE) [27], have been effectively applied [28, 29]. In this paper, we demonstrate that the information-theoretic approach can be successfully applied for soft robots. In particular, when the soft body is actuated, it can propagate the effect of this actuation through the entire body with a certain amount of time delay. We specifically focus on capturing this delayed interaction regime intrinsic to soft body dynamics. We use a recently proposed measure called *momentary information transfer* (MIT) [24], which is particularly suited to detect delayed couplings among time series, and introduce several variations of this measure to cope with practical requirements when applied to soft robotic platforms. Using a physical soft robotic platform, we demonstrate the power of our proposed measures from two aspects. First, we infer the delayed interaction regime as a strength of information transfer from a statistical point of view. This aspect is useful for evaluating an intrinsic property corresponding to the soft robotic platform in use. Second, we monitor and visualize how the information propagates through the body dynamically. This scheme is beneficial for detecting how external damage in a spatiotemporal point spreads through the entire body, which would be difficult to characterize only by looking at the behavior and the dynamics of the body.

This paper is organized as follows. In the following section, we present the measure MIT and discuss why this measure can perform better than the conventional TE in inferring delayed information transfer. Then, we modify this measure systematically to meet specific requirements in real-world applications. Here, we propose several variants of MIT, which will be used for our analyses. Next, we introduce a soft robotic platform inspired by the octopus and explain the experimental procedures to apply these measures. We demonstrate the effectiveness of our proposed scheme in revealing the delayed information transfer in the soft robotic platform. Finally, we give concluding remarks, including a possible extension scenario of our approach for general applications to soft robots, and discuss future work.

2. Methods

2.1. Information theoretic approach for delayed information transfer in spatiotemporal dynamics

In this section, we begin with a brief overview of the information-theoretic concepts required in this paper. The Shannon entropy [30] is one of the basic quantities in information theory, which defines the average uncertainty associated with the state x of a random variable X as

$$H(X) = -\sum_{x} p(x) \log_2 p(x), \tag{1}$$

where $p(x)$ denotes a probability distribution of X. The base of the logarithm is taken as 2 throughout this paper. Thus, the unit of all measures presented here is unified as

bit. The MI between two variables X and Y measures the average information gained about X by knowing Y, or vice versa, as follows:

$$M_{XY} = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}, \tag{2}$$

where $p(x,y)$ is a joint probability distribution of variables X and Y [30]. For statistically independent distributions, $p(x,y) = p(x)p(y)$ and $M_{XY} = 0$. If statistical dependencies exist, $M_{XY} > 0$. MI is a fundamental measure in information theory and used to evaluate an association between two or more variables, which naturally encompass linear and nonlinear dependencies. However, as can be seen from the equation, MI is intrinsically symmetric under the exchange of the two variables X and Y, which means that MI does not contain any directional information.

In this paper, we focus on the notion of information transfer, which requires capturing both directional and dynamic relations from an information source to a receiver. Therefore, MI is insufficient for this purpose. There are several measures proposed to address these inadequacies. A well-known measure is TE, proposed by Schreiber [27]. TE is a measure of the information transfer from the driving system (Y) to the responding system (X). Let us write x_t and y_t for the values of two temporal processes X_t and Y_t, respectively. TE essentially quantifies the deviation from the generalized Markov property: $p(x_{t+1}|x_t^{(K)}) = p(x_{t+1}|x_t^{(K)}, y_t^{(L)})$, where $p(x_{t+1}|x_t^{(K)})$ denotes a transition probability from $x_t^{(K)}$ to x_{t+1}, and K and L are the length of the embedding vectors. If the deviation from a generalized Markov process is small, then the state $y_t^{(L)}$ can be assumed to have little relevance to the transition probabilities from $x_t^{(K)}$ to x_{t+1}. If the deviation is large, however, then the assumption of a Markov process is not valid. The incorrectness of the assumption can be expressed by the TE, formulated as a specific version of the Kullback-Leibler entropy [27, 30]:

$$TE_{Y\to X} = \sum_{x_{t+1},x_t^{(K)},y_t^{(L)}} p(x_{t+1}, x_t^{(K)}, y_t^{(L)}) \log \frac{p(x_{t+1}|x_t^{(K)}, y_t^{(L)})}{p(x_{t+1}|x_t^{(K)})}, \tag{3}$$

where the index $TE_{Y\to X}$ indicates the influence of Y on X and can thus be used to detect the directed information transfer. In other words, TE measures how well we can predict the transition of system X by knowing system Y. TE is non-negative, and any information transfer between the two variables results in $TE_{Y\to X} > 0$. If the state $y_t^{(L)}$ has no influence on the transition probabilities from $x_t^{(K)}$ to x_{t+1}, or if the two time series are completely synchronized, then $TE_{Y\to X} = 0$. By introducing a time delay τ, TE has been used to infer delayed couplings such as

$$TE_{Y\to X}(\tau) = \sum_{x_{t+\tau},x_t^{(K)},y_t^{(L)}} p(x_{t+\tau}, x_t^{(K)}, y_t^{(L)}) \log \frac{p(x_{t+\tau}|x_t^{(K)}, y_t^{(L)})}{p(x_{t+\tau}|x_t^{(K)})}, \tag{4}$$

where the index $TE_{Y\to X}$ indicates the influence of Y on X with a delay τ ($\tau = 1, 2, ...$).

A weak point of TE is that it fails at detecting delayed couplings [24, 31]. One attempt to overcome this drawback, and to capture delayed information transfer, has

been presented in [24], where the authors proposed the measure MIT, defined as

$$MIT_{Y\to X}(\tau) = \sum_{x_{t+\tau}^{(K+1)}, y_t^{(L)}} p(x_{t+\tau}^{(K+1)}, y_t^{(L)}) \log \frac{p(x_{t+\tau}|x_{t+\tau-1}^{(K)}, y_t^{(L)})}{p(x_{t+\tau}|x_{t+\tau-1}^{(K)}, y_{t-1}^{(L-1)})}, \tag{5}$$

where the index $MIT_{Y\to X}(\tau)$ indicates the MIT from y_t to $x_{t+\tau}$ and τ ($\tau = 1, 2, ...$) denotes a time delay. Note that, in the equation, $(x_{t+\tau}, x_{t+\tau-1}^{(K)})$ and $(y_t, y_{t-1}^{(L-1)})$ are unified into $x_{t+\tau}^{(K+1)}$ and $y_t^{(L)}$, respectively, for simplicity. It is obvious from the equation that MIT represents the information transfer from y_t to $x_{t+\tau}$ under the condition of the joint past of y_t and $x_{t+\tau}$. This specific type of conditioning enables MIT to have a better resolution for detecting point-to-point information transfer in a temporal dimension. Similar considerations have also been presented in [32, 33]. Further details on the performance of MIT in a real-world time series, and on the comparisons with TE are given in [24].

Our aim in this paper is to apply information-theoretic measures to spatiotemporal time series in a physical soft robot platform. In particular, we are interested in capturing the delayed information transfer through soft body dynamics and characterizing how it propagates through the entire body. A number of useful methods for this purpose have been proposed in the literature in different contexts. Based on MIT, we will systematically place specific conditions for our requirements and explain how to incorporate these methods and obtain feasible measures for our purpose. This will help us to understand how each extension clearly meets our requirement and assist in applying our approach to any soft robotic platform in use. A rough sketch of our requirements and modification strategies of the measure are given next, which will be explained in detail in the following sections.

The first requirement is for ease in handling the measure for a real-world time series. As previously explained, the information-theoretic measures are functionals of probability distributions, and there are variations to methods proposed to estimate the distributions from an obtained time series. In a real-world time series, this procedure often requires careful preprocessing of the data, setting of the distribution estimators, and fine-tuning their parameters, due to environmental and observational noises. Furthermore, the obtained results are usually not easy to reproduce without specifying the details of these methods. One of the well-known treatments for these issues is to introduce a class of the probability distribution estimator based on permutation partitioning for the values of the data. The corresponding uncertainty measure, *permutation entropy* (PE), is introduced in [34] and quantifies the uncertainty of the local orderings of values, unlike the usual entropy, which quantifies the uncertainty of the values themselves. This approach enables a natural discretization for time series data that does not require any knowledge about the range of values of the time series beforehand. It can be applied to raw data without further model assumptions, because the permutation orderings refer only the local and neighboring values in time series data. The information-theoretic measures based on the permutation partitioning have been demonstrated to be easy to implement relative to the other traditional methods

[34], fast to compute [35], robust against noise [34, 36, 37, 38, 39], and applicable to non-stationary and transient time series data [40], which meets our requirement.

The second requirement is to make the measure applicable to a multivariate time series. Body dynamics generated by soft robotic platforms often reveal high-dimensional spatiotemporal dynamics compared with rigid ones. As we saw for MIT, measuring delayed couplings in high resolution needs special care in conditioning the past time series, both for the information source and receiver. We will explain how to deal with this conditioning for multivariate time series data in detail, specifically for data obtained from the physical platform.

The final requirement is to make it possible for the measure to characterize a local information transfer profile within the spatiotemporal dynamics. It is valuable to track and visualize how the information dynamically propagates through the entire body, along with its corresponding body dynamics. However, just like other information-theoretic measures, due to their statistical nature, MIT only provides an expectation value of the amount of information transferred that originates from the global average, which does not directly correspond to the body dynamics themselves. We will explain in detail, a method to make MIT applicable for our purpose by localizing it to each spatiotemporal point of the body dynamics.

2.1.1. Momentary sorting information transfer: Permutation analogue of MIT As previously explained, PE quantifies the uncertainty of the local orderings of values, instead of the uncertainty of the values themselves. Let $x_t^{(L)}$ represent an L-dimensional embedding vector from the obtained time series x'_t, and $\hat{x}_t^{(L)}$ be a sequence of numbers representing the orderings of $x_t^{(L)}$. Based on the permutations of the values, $\hat{x}_t^{(L)}$ is generated as follows: $x_t^{(L)} = (x'_t, x'_{t-1}, ..., x'_{t-(L-1)})$, and the values are arranged in ascending order, $x'_{t-(o_t(1)-1)} \leq x'_{t-(o_t(2)-1)} \leq ... \leq x'_{t-(o_t(L)-1)}$. A symbol is thus defined as $\hat{x}_t^{(L)} \equiv (o_t(1), o_t(2), ..., o_t(L)) \in \hat{X}_t$, where $\hat{X}_t$ is the set of symbols generated in the temporal process X_t. Based on the generated symbols $\hat{x}_t^{(L)}$, PE is expressed as:

$$H(\hat{X}_t) = -\sum_{\hat{x}_t^{(L)}} p(\hat{x}_t^{(L)}) \log p(\hat{x}_t^{(L)}), \tag{6}$$

where $p(\hat{x}_t^{(L)})$ is the probability of the occurrence of $\hat{x}_t^{(L)}$ in the set of symbols $\hat{X}_t$. In spite of the differences between the procedures, it was proven that the PE rate is equal to the usual entropy rate for some conditions [41, 42, 43, 44, 45].

A number of permutation analogues for the information-theoretic measures have been proposed and the relations with their original measure have been investigated (e.g., systematic investigations on Kolmogorov-Sinai entropy can be found in [46, 47, 48, 49, 50, 51] and investigations on TE and its related measures can be found in [52, 53, 54]). Similar to MIT, its permutation version, proposed in [24], is called *momentary sorting information transfer* (MSIT), expressed as:

$$MSIT_{Y\to X}(\tau) = \sum_{\hat{x}_{t+\tau}^{(K+1)}, \hat{y}_t^{(L)}} p(\hat{x}_{t+\tau}^{(K+1)}, \hat{y}_t^{(L)}) \log \frac{p(\hat{x}_{t+\tau} | \hat{x}_{t+\tau-1}^{(K)}, \hat{y}_t^{(L)})}{p(\hat{x}_{t+\tau} | \hat{x}_{t+\tau-1}^{(K)}, \hat{y}_{t-1}^{(L-1)})}, \tag{7}$$

where, similar with MIT, the index $MSIT_{Y\to X}(\tau)$ indicates the MSIT of $\hat{y}_t$ to $\hat{x}_{t+\tau}$ and τ ($\tau = 1, 2, ...$) denotes a time delay. Note that for $\hat{x}_{t+\tau}$, the permutation orderings among $\hat{x}^{(K+1)}_{t+\tau}$ are considered, because when we calculate the measure, we first obtain the joint entropies from the corresponding probability distributions, namely $p(\hat{x}^{(K+1)}_{t+\tau}, \hat{y}^{(L)}_t)$ and $p(\hat{x}^{(K+1)}_{t+\tau}, \hat{y}^{(L-1)}_{t-1})$, with the obtained time series. A similar treatment has been also adopted in [55].

In the next section, we will show how MSIT can be extended for the case of multivariate time series.

2.1.2. MSIT for multivariate time series As we saw in the case for M(S)IT, it is important to employ a conditioning for the information receiver and sender in focus. This guarantees to yield a high-resolution measure for the bivariate time series. When we extend the measure for the multivariate case, we have to take particular care of the conditioning, because other possible causal information contributors—apart from the information source under consideration—may be involved in the time series as well. It is therefore important to "filter out" the effect of these other information sources, which might otherwise occlude effect of the source in question. There have been several studies to address these issues. In [25, 26], the generalized MIT has been proposed, a method to infer all causal contributors to the receiver, based on the graphical model approach. Their approach is general, because the method allows reconstruction of an interaction modality of the multivariate data in terms of a process graph. This graph is built from scratch without having any knowledge about the dynamical property of the underlying system in focus (such as network topology of the system).

In this paper, we will adopt their approach only partially. We take into account a physical constraint of the soft robotic platform, as a support for estimating the other causal information contributors, apart from the information source. We first assume that the experimenter knows where each time series comes from in space, namely the location in the physical body. We further assume that the degree of information transfer depends on the relative spatial and temporal distance; if the information source and receiver are more distant and delayed, then the strength of the information transfer decreases, which would be a natural assumption, especially when the information transfer substrates are mediated with physical materials. Based on these assumptions, we introduce a simple one-dimensional space as an example, where each point in space is expressed as cell i ($i = 1, 2, ..., N$), resulting in N time series. Accordingly, the senders and receivers are denoted as spatially ordered cells, the MSIT for spatiotemporal time series from cell i to cell j can be expressed as

$$MSIT_{i\to j}(\tau) = \sum_{\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}_{i,j,t}} p(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}_{i,j,t}) \log \frac{p(\hat{x}_{j,t+\tau}|\hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t}, \hat{v}_{i,j,t})}{p(\hat{x}_{j,t+\tau}|\hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L-1)}_{i,t-1}, \hat{v}_{i,j,t})}, \quad (8)$$

$$\hat{v}_{i,j,t} = \{\hat{x}^{(M)}_{q,t+\tau'} | \forall(q, \tau'), 1 \le q \le N, \tau' < \tau \ (except \ \hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t})\}, \quad (9)$$

where $\hat{v}_{i,j,t}$ denotes possible information sources to cell j excluding $\hat{x}^{(K+1)}_{j,t+\tau}$ and $\hat{x}^{(L)}_{i,t}$, and M, q, and τ' denote their embedding dimension, cell number, and a time delay, respectively. One could adjust these parameters for each information source; for example, if there are overlapping sources in a temporal dimension, such as $(\hat{x}^{(M)}_{q,t+1}, \hat{x}^{(M)}_{q,t})$, we can combine them into one as $\hat{x}^{(M+1)}_{q,t+1}$. Note that it is possible to condition information sources for $\hat{x}^{(L)}_{i,t}$ to sharpen the point-to-point MIT, but in this paper, we only consider $\hat{x}_{i,t}$ under the condition of its direct past $\hat{x}^{(L-1)}_{i,t-1}$ for simplicity. Now, we specify the conditions of Equations (8) and (9) for our spatiotemporal constraint, expressed as

$$MSIT^{ST}_{i\to j}(\tau) = \sum_{\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t}} p(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t}) \log \frac{p(\hat{x}_{j,t+\tau} | \hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})}{p(\hat{x}_{j,t+\tau} | \hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L-1)}_{i,t-1}, \hat{v}^{ST}_{i,j,t})}, \quad (10)$$

$$\hat{v}^{ST}_{i,j,t} = \{\hat{x}^{(M)}_{j\pm r, t+\tau-\tau_r} | \forall (r, \tau_r), 1 \le r \le N_r, 1 \le \tau_r \le T_r \ (except\ \hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t})\}, \quad (11)$$

where $MSIT^{ST}_{i\to j}(\tau)$ and $\hat{v}^{ST}_{i,j,t}$ represent the MIT from i to j with delay τ and the possible information sources to cell j under the spatiotemporal constraint, respectively, parameters r and τ_r determine a cell and a time delay for possible information sources to cell j excluding $\hat{x}^{(K+1)}_{j,t+\tau}$ and $\hat{x}^{(L)}_{i,t}$, and N_r and T_r specify the spatial and temporal range around the information receiver, respectively; if r and τ_r get larger, then the possible information source is more spatiotemporally distant from the receiver. Figure 1 visualizes the $MSIT^{ST}_{i\to j}(\tau)$ and summarizes the effects of the different parameters involved.

Note that $MSIT^{ST}_{i\to j}(\tau)$ can be also denoted as a sum of permutation joint entropies:

$$MSIT^{ST}_{i\to j}(\tau) = -H(\hat{X}^j_{t+\tau}, \hat{X}^i_t, \hat{V}^{i,j}_t) + H(\hat{X}^j_{t+\tau-1}, \hat{X}^i_t, \hat{V}^{i,j}_t) \quad (12)$$

$$-H(\hat{X}^j_{t+\tau}, \hat{X}^i_{t-1}, \hat{V}^{i,j}_t) + H(\hat{X}^j_{t+\tau-1}, \hat{X}^i_{t-1}, \hat{V}^{i,j}_t), \quad (13)$$

where $\hat{X}^j_{t+\tau}$, $\hat{X}^j_{t+\tau-1}$, $\hat{X}^i_t$, $\hat{X}^i_{t-1}$, and $\hat{V}^{i,j}_t$ are the temporal processes for the values of $\hat{x}^{(K+1)}_{j,t+\tau}$, $\hat{x}^{(K+1)}_{j,t+\tau-1}$, $\hat{x}^{(L)}_{i,t}$, $\hat{x}^{(L)}_{i,t-1}$, and $v^{ST}_{i,j,t}$, respectively, and $H(\hat{X}^j_{t+\tau}, \hat{X}^i_t, \hat{V}^{i,j}_t)$, $H(\hat{X}^j_{t+\tau-1}, \hat{X}^i_t, \hat{V}^{i,j}_t)$, $H(\hat{X}^j_{t+\tau}, \hat{X}^i_{t-1}, \hat{V}^{i,j}_t)$, and $H(\hat{X}^j_{t+\tau-1}, \hat{X}^i_{t-1}, \hat{V}^{i,j}_t)$ are joint entropies of corresponding temporal processes. When we calculate the value of $MSIT^{ST}_{i\to j}(\tau)$, we first calculate the joint and single entropies on the right side of this equation from the corresponding probability distributions, namely $p(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})$, $p(\hat{x}^{(K+1)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})$, $p(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t-1}, \hat{v}^{ST}_{i,j,t})$, and $p(\hat{x}^{(K+1)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t-1}, \hat{v}^{ST}_{i,j,t})$, with the obtained time series.

2.1.3. Localizing MSIT for spatiotemporal systems In this section, we present the procedure for making MSIT local. This procedure makes it possible to reveal an information transfer profile for each spatiotemporal point by directly corresponding the measure to the observed time series, which would be especially useful for monitoring the interaction modality of the local dynamics from the information-theoretic point of view. We keep carrying on the physical constraint introduced in the previous section and explain the procedure based on the measure in Equations (10) and (11).

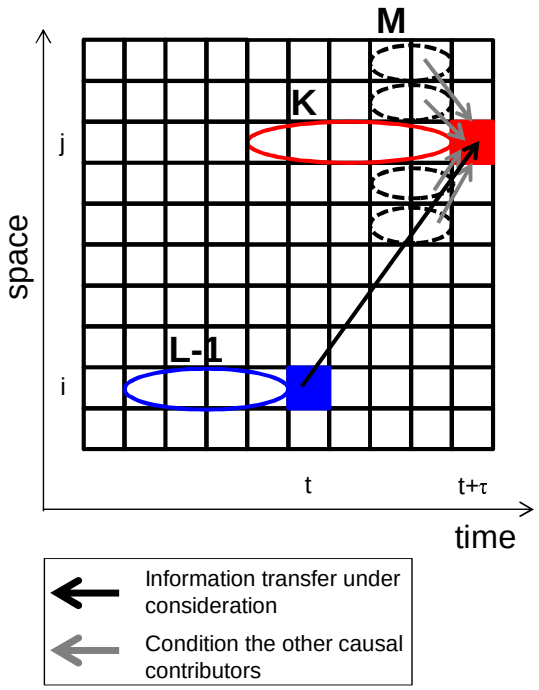


Figure 1. Schematic expressing the MSIT in a spatiotemporal system. The black arrow shows the information transfer under consideration and the gray arrow shows the other causal contributors to the destination cell. The MSIT for the case when $(\tau, K, L, M, N_r, T_r) = (4, 5, 5, 2, 2, 1)$ is shown.

In [56], Lizier *et al.* focused on the fact that, in calculating the information-theoretic measure from the experimental data, the associated probability $p(x)$ is operationally equivalent to the ratio of the count of observations $c(x)$ of states, to the total number of observations O made. In applications to time series data for MSIT, the number of observations is finite, and $p(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})$ in Equation (10) can be expressed as $p(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t}) = c(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})/O$. Then, MSIT in Equation (10) can be expressed as follows:

$$MSIT^{ST}_{i\to j}(\tau) = \frac{1}{O} \sum_{\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t}} \left(\sum_{a=1}^{c(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})} 1 \right) \log \frac{p(\hat{x}_{j,t+\tau} | \hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})}{p(\hat{x}_{j,t+\tau} | \hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L-1)}_{i,t-1}, \hat{v}^{ST}_{i,j,t})}. \quad (14)$$

By considering that a double sum running over each actual observation a for each possible tuple observation $(\hat{x}^{(K+1)}_{j,t+\tau}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})$ is nothing but a single sum over all O observations, we obtain the following:

$$MSIT^{ST}_{i\to j}(\tau) = \frac{1}{O} \sum_{all\ \ observations} \log \frac{p(\hat{x}_{j,t+\tau} | \hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})}{p(\hat{x}_{j,t+\tau} | \hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L-1)}_{i,t-1}, \hat{v}^{ST}_{i,j,t})}. \quad (15)$$

Thus, we can write MSIT as the global average over *local momentary sorting information transfer* (LMSIT), $LMSIT^{ST}_{i\to j}(\tau)$, defined as,

$$MSIT^{ST}_{i\to j}(\tau) = \langle LMSIT^{ST}_{i\to j}(\tau) \rangle, \quad (16)$$

$$LMSIT^{ST}_{i\to j}(\tau) = \log \frac{p(\hat{x}_{j,t+\tau}|\hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L)}_{i,t}, \hat{v}^{ST}_{i,j,t})}{p(\hat{x}_{j,t+\tau}|\hat{x}^{(K)}_{j,t+\tau-1}, \hat{x}^{(L-1)}_{i,t-1}, \hat{v}^{ST}_{i,j,t})}, \tag{17}$$

where $\langle X \rangle$ denotes the expectation value of X. The term $\hat{v}^{ST}_{i,j,t}$ takes the same expression as Equation (11). Note that LMSIT can have a negative value. The negative value of LMSIT means that the sender is misleading about the prediction of the receiver's next state [56]. This measure is defined for every spatiotemporal receiver $(j, t+\tau)$, forming a spatiotemporal profile for every information source i in focus. Thus, in a practical sense, as we see later, it would be more reasonable to adjust the time index for the information receiver as (j, t). This can be easily done by simply shifting the time index without altering the important factor of the measure as

$$LMSIT^{ST}_{i\to j}(\tau) = \log \frac{p(\hat{x}_{j,t}|\hat{x}^{(K)}_{j,t-1}, \hat{x}^{(L)}_{i,t-\tau}, \hat{v}^{ST}_{i,j,t-\tau})}{p(\hat{x}_{j,t}|\hat{x}^{(K)}_{j,t-1}, \hat{x}^{(L-1)}_{i,t-\tau-1}, \hat{v}^{ST}_{i,j,t-\tau})}, \tag{18}$$

$$\hat{v}^{ST}_{i,j,t-\tau} = \{\hat{x}^{(M)}_{j\pm r,t-\tau_r} | \forall(r,\tau_r), 1 \le r \le N_r, 1 \le \tau_r \le T_r \ (except\ \hat{x}^{(K+1)}_{j,t}, \hat{x}^{(L)}_{i,t-\tau})\}. \tag{19}$$

This forms the expression of LMSIT addressed throughout this paper. Similar local measures have been proposed in our earlier works [39, 57], which are based on the permutation version of TE called *symbolic transfer entropy* (STE) [35]. Our proposed measures extend our previous measures in terms of the condition of the information sender, which is expected to sharpen resolution, especially for detecting delayed information transfer. As in the case with $MSIT^{ST}_{i\to j}(\tau)$, to calculate the value of $LMSIT^{ST}_{i\to j}(\tau)$, we calculate the joint and single probability distributions, namely $p(\hat{x}^{(K+1)}_{j,t}, \hat{x}^{(L)}_{i,t-\tau}, \hat{v}^{ST}_{i,j,t-\tau})$, $p(\hat{x}^{(K+1)}_{j,t-1}, \hat{x}^{(L)}_{i,t-\tau}, \hat{v}^{ST}_{i,j,t-\tau})$, $p(\hat{x}^{(K+1)}_{j,t}, \hat{x}^{(L)}_{i,t-\tau-1}, \hat{v}^{ST}_{i,j,t-\tau})$, and $p(\hat{x}^{(K+1)}_{j,t-1}, \hat{x}^{(L)}_{i,t-\tau-1}, \hat{v}^{ST}_{i,j,t-\tau})$, with the obtained time series.

In this section, we introduced MSIT and LMSIT with spatiotemporal constraints to meet our requirements. In the next section, we will explain our physical platform with soft robotic arm, and by using the body dynamics generated by the arm, we illustrate how these measures can capture the characteristic properties of soft body dynamics, in particular, the delayed information transfer, and demonstrate the power of the measures.

2.2. Experimental setup

In this section, our experimental platform, equipped with a soft robotic arm, is presented and the experimental procedures are explained.

2.2.1. Platform The experimental platform consists of a soft robotic arm, its actuation, sensing, control systems, and a water tank containing fresh water as the underwater environment (Fig. 2 (a)). There are several soft robotic arms proposed in the literature (see, e.g., [12, 14, 11, 13]). Our soft robotic arm is based on the arm proposed in [12], which mimics the morphology and material characteristics of the octopus arm, and is made of commercially available silicone rubber (ECOFLEX$^{\text{TM}}$00-30). The arm

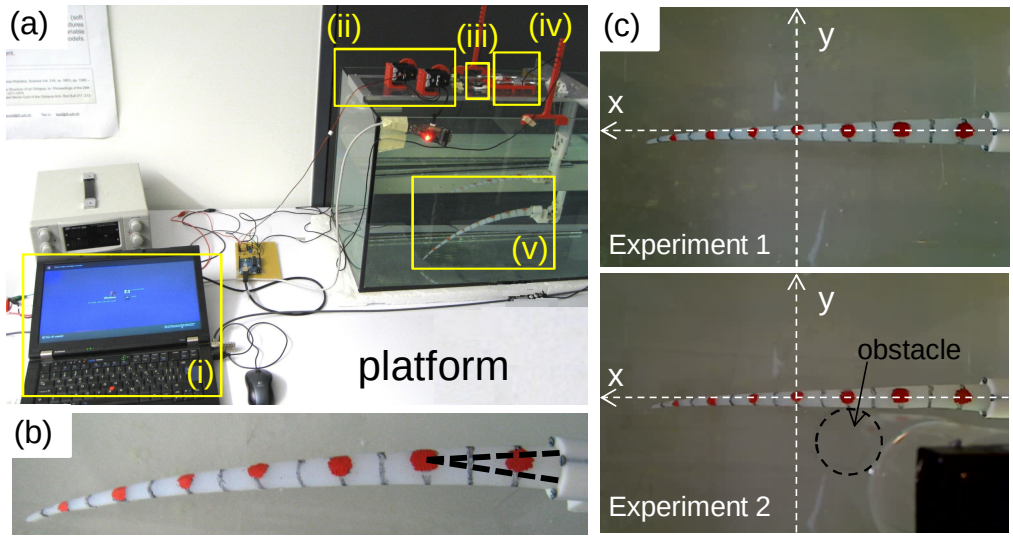


Figure 2. Experimental setups. (a) An overall experimental platform. It consists of a laptop PC (i), two servo motors (ii), one video camera (iii), two force sensors (iv), and a soft robotic arm (v). (b) Close-up image of the soft robotic arm used in this paper. Dashed lines represent the cables embedded in the arm. Red dots and black lines on the arm are references for visual tracking. (c) Initial arm configurations of Experiment 1 (upper figure) and Experiment 2 with the obstacle (lower figure). In each experiment, the arm is set to the relaxed position. In Experiment 2 with the obstacle, the obstacle is made of a transparent plastic tube on a black mount and is set near the R3 of the arm to affect its behavior. The obstacle is fixed so that the arm motion cannot change the position of the obstacle. The base point of the x-y coordinates for the tracking is set to near R4 of the arm in both experiments.

embeds two nonextensible fishing cables symmetrically to the center of the arm, as shown by the dashed lines in Fig. 2 (b). By pulling these two cables via servo motors (Dynamixel^{TM}AX-12A+), the arm is actuated (Fig. 2 (a)). The total length of the cone-shaped soft arm is 310 mm. The active part of the arm is 80 mm and the remaining 230 mm part of the arm is completely passive. Two force sensors measure the tension of the cables. The force sensor signals are amplified and sent to a PC serial port through an Arduino^{TM}UNO board, with ADC outputs integer values between 0 and 1024, which correspond linearly to forces of 0 to 10 N (Fig. 2 (a)). A camera (Logitech^{TM}Webcame Pro 9000) is placed on the top of the platform to record the motion of the soft silicone arm (Fig. 2 (a)).

A java program running on a laptop PC sends the motor commands to the motors. The unit representing "time" in this paper is a "timestep" based on the sensing and actuation loop in the program (this is about 0.03 s in a physical time). For each timestep, a motor command is sent to each motor and the cable tensions are recorded. According to the motor command, the servo motor angle is adjusted. In this paper, motor commands are set as binary values, $M = \{+1, -1\}$. If the command gives $+1$ or -1, then the motor is controlled to move from the current position toward the maximum

position (L_{max}) or the relaxed position (L_{relax}), respectively. For each motor, L_{max} was determined to avoid the tip of the arm touching the water tank while moving. Note that, in the current setting, the motor command does not always take the roller position to L_{max} or L_{relax}, but rather decides the direction in which to move for each timestep. If the command gives $+1$ or -1 when the current position is in L_{max} or L_{relax}, respectively, then the position will not change from the current position.

For the analyses, we collected the time series of the two motor commands (M1 and M2) and two force sensors (f1 and f2) for each timestep and tracked the positions of the reference points on the arm (R) in the x-y plane, by using the tracker proposed in [58] (Fig. 2 (b) and Fig. 3). The tracked time series were manually aligned to the motor command time series after each run. For the analyses in this paper, we used only 6 reference points, which are indexed from R1 to R6, from the base to the tip. The nearest reference point to the tip was too fast to allow stable tracking of the motion and was not used for the analyses. Based on these 6 reference points for each timestep, we fitted the points with a 5th order polynomial curve and for each interval between two neighboring reference points, we segmented the curve into 20 equidistant fragments, resulting in 101 endpoints for fragments, including the reference points (Fig. 3, and see detailed procedures in Appendix). The dynamics of these 101 points are used as a representation of the arm's body dynamics (S), which are indexed from S_1 to S_{101} from the base to the tip. For the body dynamics S, we used only the y-coordinate for the information-theoretic analysis, which reflects the motion of the arm better than the x-coordinate (Fig. 2 (c)).

2.2.2. Experimental procedures We conducted two groups of experiments, where all the experiments were based on open-loop control. Experiment 1 was designed to reveal the informational structure intrinsic to the soft robotic arm in our platform setting. As explained in the previous section, the arm is controlled by two cables embedded in the arm. First, we investigated how the effects of these actuations transfer through the entire body. Second, we investigated how each body part transfers the information to the other parts of the body. These two investigations employ the MSIT, which measures delayed couplings. In this experiment, to avoid correlations induced by coordinated motor commands, we sent random motor commands to each motor for each timestep. By doing this, we moved each roller in random positions from L_{relax} to L_{max} over time, and effectively detect the coupling regime intrinsic to the setting. We ran the experiment for 5 trials with 5000 timesteps each, starting from the same initial arm configuration (Fig. 2 (c)).

In Experiment 2, we investigated a local information transfer profile in the spatiotemporal dynamics of our soft robotic arm. We set a simple oscillatory motor command and observed the behavior with an obstacle in the water tank (Fig. 2 (c)). In this experiment, we aimed to capture how the environmental impact spreads dynamically through the entire body. This is difficult to reveal directly with a measure that deals with a global average of information transfer, because the effect of damage or contact to

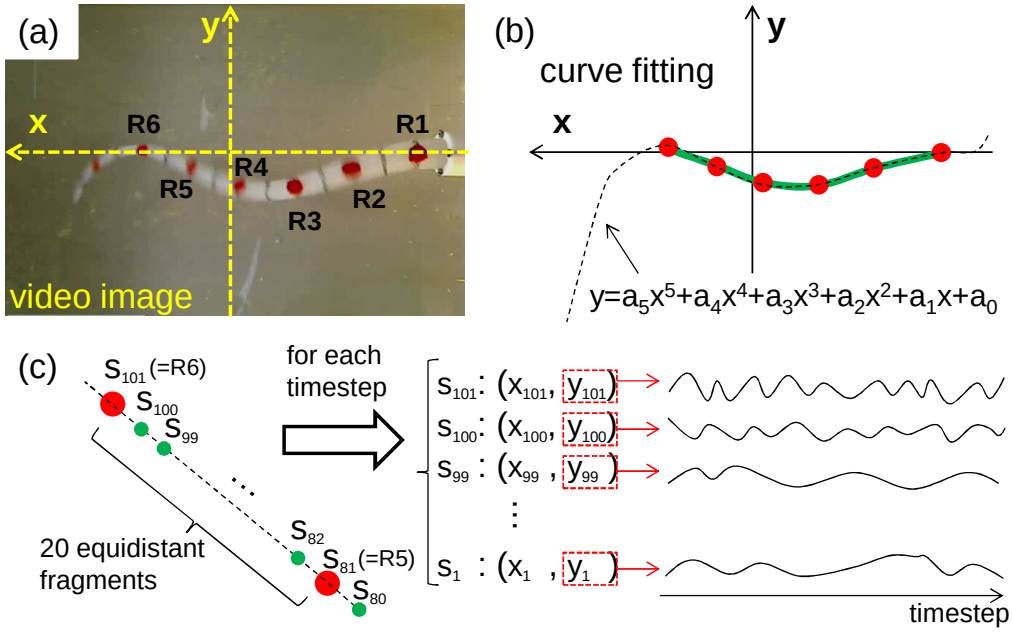


Figure 3. Schematics showing the procedures to prepare the time series representing the soft body dynamics. (a) For each timestep, the (x, y)-coordinates for 6 reference points are tracked using the tracking system [58]. Note that the nearest reference point to the tip is not used for this analysis (see the main text for details). (b) The 5th order polynomial curve fitting the body posture based on the obtained coordinates of the reference points. (c) Close-up of the curve between R6 and R5 as an example for the preparation of S. For each neighboring reference point, the curve is segmented into 20 equidistant fragments, and the endpoints of the fragments are indexed from S_1 to S_{101} from the base to the tip, including the reference points, and their (x, y)-coordinates are collected for each timestep. The dynamics of the y-coordinates of these endpoints are the representation of the body dynamics in this paper.

the body is usually not a constant event, but rather one that occurs at a specific point of time and space. We placed a round plastic tube (obstacle) next to the arm, which was set to affect the behavior of the arm (Fig. 2 (c)). For the motor commands, +1 and -1 were set for T timesteps in alternate shifts in an opposite phase each. The time window T was set to the maximum timesteps needed for the motor to rotate from the relaxed position to the maximum one, which, in our setting, was 10 timesteps. Accordingly, one cycle resulted in 20 timesteps. We ran the system for 5000 timesteps, which were 250 cycles, from the initial arm configuration depicted in Fig. 2 (c). In the experiment, we also ran the system with random motor commands and the same oscillatory motor command without the obstacle and compared the results.

Note that, for both experiments, the experimental condition is determined to make the time series data stationary, and the probability distributions of the variables required to calculate measures, MSIT and LMSIT, are estimated via relative frequencies obtained from the data. Accordingly, in each experimental trial for both experiments, first 100 timesteps are discarded to avoid the initial transients in the time series data for the

analyses. This procedure is remained fixed for all the experimental trials throughout this paper.

Preliminary investigations of information-theoretic analyses on a soft robotic platform have been performed in [57, 59] without considering the effect of delayed information transfer. In these works, the analyses were conducted based on STE. However, as explained earlier, STE has drawbacks in detecting delayed information transfer. In this paper, by introducing MSIT, we have largely extended our previous approach.

3. Results

In this section we present the results for both Experiment 1 and 2 in detail. As explained in the previous section, Experiment 1 is focused on revealing the delayed information transfer intrinsic to the soft robotic arm. For the analysis of this experiment, we make use of MSIT expressed in Equation (10) and (11) (that is, $MSIT^{ST}_{i\to j}(\tau)$). In Experiment 2, unlike the analysis in Experiment 1, we aim to reveal how the information propagates dynamically through the body by visualizing the local information transfer in each spatiotemporal point. This analysis is conducted by using LMSIT expressed in Equation (18) and (19) (that is, $LMSIT^{ST}_{i\to j}(\tau)$). Throughout our analyses, the embedding dimensions of the measures are set as $(K, L, M) = (2, 3, 2)$. As explained in the earlier sections, to calculate the measures we should first estimate the joint and single probability distributions from the obtained time series. To get a reliable estimate within the limitation of the finite data set, it is preferable to keep the embedding dimensions relatively short [24].

3.1. Experiment 1

In this experiment, we see how MSIT can quantitatively characterize the information-theoretic structure intrinsic to the soft robotic arm. For this purpose, we adopted a random motor command for the actuation of the system to avoid additional correlations provided by the external. Figure 4 shows the example of the time series of the random motor commands $(M1(t), M2(t))$, the corresponding force sensory values $(f1(t), f2(t), f1(t) - f2(t))$, and the y-coordinates of the body parts S_i. By using these obtained time series, we aim to analyze the following: (I) information transfers from the time series of the cable tension to that of each body part ($MSIT^{ST}_{f\to S_i}(\tau)$) and (II) information transfers between the time series of each body part ($MSIT^{ST}_{S_i\to S_j}(\tau)$). In (I), since the cable tensions are driven by the motor command, this analysis is to see how the external actuation to the system transfers the information through the body. For the analysis, we used the time series $f(t) = f1(t) - f2(t)$ for the force sensors, and calculated $MSIT^{ST}_{f\to S_i}(\tau)$ for each body part i and varying delay τ. In (II), we can characterize how one body part transfers information to another. For the analysis, we calculated $MSIT^{ST}_{S_i\to S_j}(\tau)$ for each pair of body parts (i, j) and varying delay τ. Note

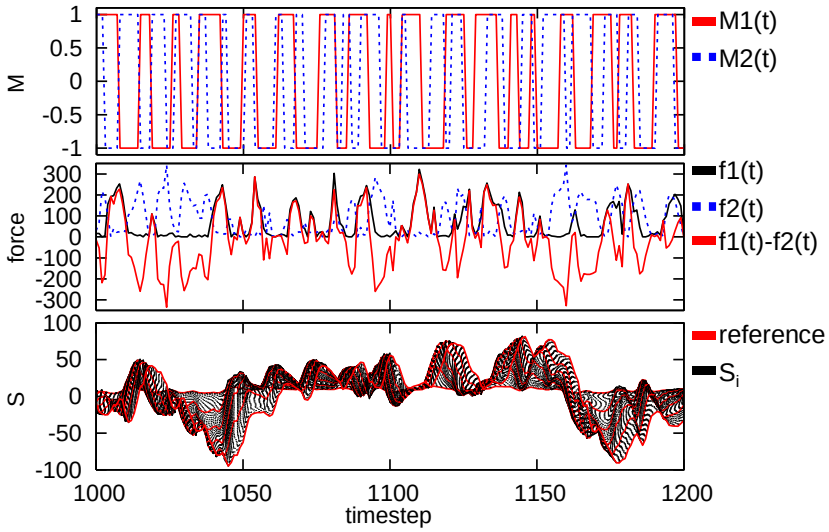


Figure 4. Typical examples of the time series for the random motor commands $(M1(t), M2(t))$ (upper plot), force sensors $(f1(t), f2(t), f1(t) - f2(t))$ (middle plot), and the body parts S_i (lower plot) in Experiment 1. For the plots of the body parts, 101 time series (from S_1 to S_{101}) are overlaid and 6 reference tracking points are expressed in red lines.

that we excluded the case when $i = j$. For both analyses (I) and (II), (N_R, T_R) is set to $(1, 1)$ throughout the analysis and, to avoid the bias due to the finite data set, for each analysis of $MSIT^{ST}_{i\to j}(\tau)$ we have also calculated the measure with temporally shuffled time series, and by iterating this procedure for 50 times and obtaining the averaged value, we have subtracted this value from the $MSIT^{ST}_{i\to j}(\tau)$ in focus. For each experimental setting, we performed this procedure for 5 trials of experimental data and used the averaged $MSIT^{ST}_{i\to j}(\tau)$ as a result. In addition, when the information sender is within the range specified by N_R and T_R, we need to take special care in defining $\hat{v}^{ST}_{i,j,t}$. Details for this procedure are given in the Appendix. Also, according to the setting of N_R and T_R, we have excluded the case where the information receiver or sender is S_1 or S_{101}.

Figure 5 (a) plots the results for the averaged $MSIT^{ST}_{f\to S_i}(\tau)$ by varying the body part i and the delay τ. We can clearly see that for body part closer to the base, more delayed MSIT is detected than for those closer to the tip. Figure 5 (b) overlays several cross sections of Fig. 5 (a) in delay and space (body part). As can be seen from the figure, in the body part near the arm base MSIT with shorter delay is dominant, while in the body part near the arm tip MSIT with relatively longer delay is dominant. In addition, the strength of MSIT tends to decrease from the base to the tip. To see this tendency in more detail, we have collected the maximum values of MSIT $(MSIT_{max})$ among delays and the delay τ_{max} that shows $MSIT_{max}$ for each body part and calculated the average for each (Fig. 5 (c)). As a result, we found that τ_{max} almost grows linearly as the body part changes from the base toward the tip and at around the tip τ_{max} took approximately 6 timesteps (Fig. 5 (c), left plot). In addition, the corresponding

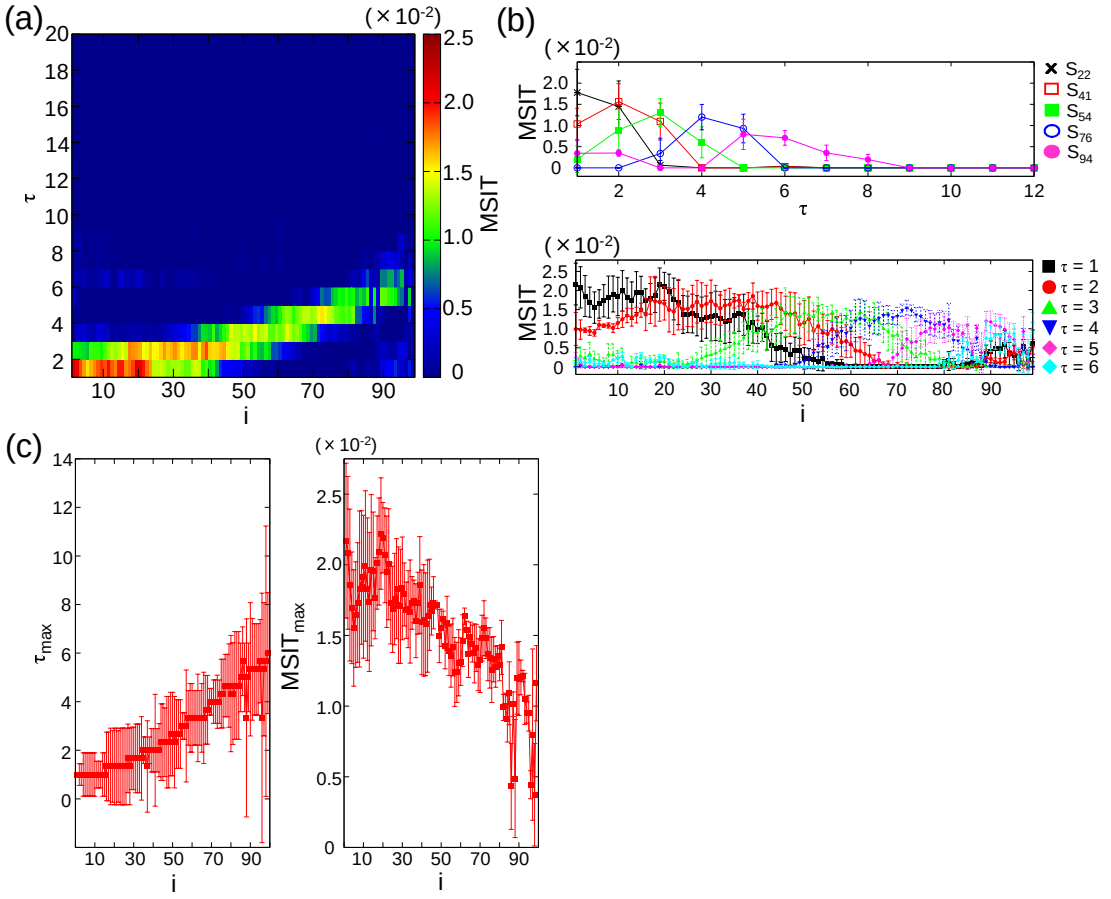


Figure 5. Results for MSIT from force sensors to the body parts. (a) Plot showing the averaged $MSIT^{ST}_{f\to S_i}(\tau)$ in i-τ plane. (b) Plots overlaying several typical cross sections of (a) in τ axis (upper figure) and in i axis (lower figure). (c) Plots showing τ_{max} (left figure) and $MSIT_{max}$ (right figure) according to each body part i. τ_{max} and $MSIT_{max}$ represent the τ, where the MSIT is maximum, and the maximum MSIT over all delays τ, respectively. For (b) and (c), the error bars show standard deviations.

$MSIT_{max}$ showed about 2.2×10^{-2} bit at around the base (S_2), while the $MSIT_{max}$ showed a lower value of about 1.2×10^{-2} bit at around the tip (S_{100}). Considering that 1 timestep is approximately 0.03 s in real physical time and the entire body is 310 mm long, this result implies that the dominant information transfer propagates with about 1.7 m/s in velocity through the body by losing the information for around 1.0×10^{-2} bit.

Next, let us see how information transfers between each body part. Here, for each information sender S_i and receiver S_j in the body parts, we calculated $MSIT^{ST}_{S_i\to S_j}(\tau)$ by varying τ from 1 to 12 and obtained the averaged MSIT ($MSIT_{average}$) as $MSIT_{average} = \frac{1}{12}\sum_{\tau=1}^{12} MSIT^{ST}_{S_i\to S_j}(\tau)$ for each trial. As in the previous experiments, we also collected the maximum MSIT ($MSIT_{max}$) among delays and the delay τ_{max} that takes $MSIT_{max}$ for each (i, j) in each trial. Figure 6 (a) shows the averaged $MSIT_{average}$, $MSIT_{max}$,

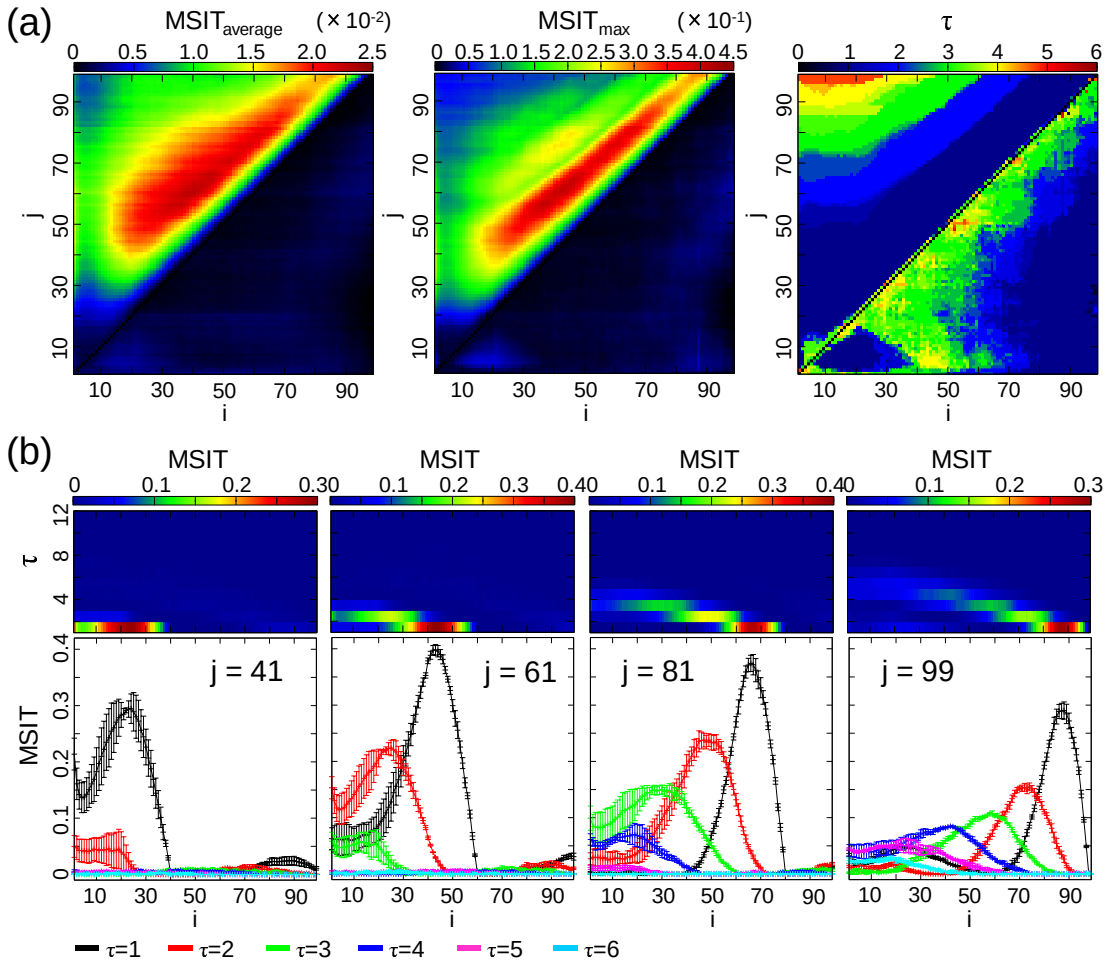


Figure 6. Results for MSIT between the body parts. (a) Plots showing the averaged $MSIT_{average}$ (left), $MSIT_{max}$ (middle), and τ_{max} (right) for each (i, j) over 5 trials. See the main text for details. (b) Plots showing the averaged $MSIT^{ST}_{f \to S_i}(\tau)$ over 5 trials according to each delay τ and the location of the information sender i by fixing the location of the the information receiver to $j = 41, 61, 81$, and 99. The error bars show standard deviations.

and τ_{max} over 5 trials for each (i, j). In the plot for the $MSIT_{average}$ and $MSIT_{max}$, we can clearly see that high MSIT values are observed only in the region of $i < j$. This suggests that the information is transferring in the direction from the base to the tip but not from the tip to the base. This result is understandable when considering the experimental conditions of our platform, and is consistent with the results from the previous experiments. Furthermore, from both plots we can confirm that, as the value for $j - i$ increases in the region of $i < j$, the values of $MSIT_{average}$ and $MSIT_{max}$ get gradually smaller. This implies that, for each body part, information strongly transfers toward the nearest neighbor in the direction from the base to the tip. For the results of τ_{max}, large τ_{max} values are observed in the region for the smaller i and the larger j. This means that the information transfer takes longer time when the body parts are

further away from each other. It is noticeable that the largest τ_{max} in this region takes approximately 6 timesteps, which is also consistent with our previous results. To see these tendencies in further detail, by fixing the information receiver to several body parts we observed the behavior of $MSIT^{ST}_{S_i \to S_j}(\tau)$ by varying the location of the information sender in the body j and the delay τ (Fig. 6 (b)). We can clearly confirm from the plot that, for each information receiver, the largest information transfers with the shortest delay from neighboring body parts located in the base side (Fig. 6 (b)).

In this experiment, we have demonstrated how the measure MSIT can be effectively used to reveal the delayed information transfer in the soft robotic arm. In the next experiment, we move on to characterize a local information transfer profile in the spatiotemporal dynamics of the soft body using LMSIT.

3.2. Experiment 2

Due to the soft flexible bodies, soft robots are sensitive to environmental/external stimulus in general. If the body receives some stimulus from the environment, the behavior of the body changes drastically and immediately. This type of event often occurs at a specific point in time and space, which makes it difficult to detect and evaluate the effect of the stimulus using statistical methods based on the global average over the entire collected time series. In those cases, it would be beneficial if we could monitor what is happening locally at each spatiotemporal point in the dynamics. In this experiment, we will see that LMSIT can be used effectively for this purpose. LMSIT is defined for each information receiver and can measure the amount of information transferred from the sender at each spatiotemporal point. Thus, the local information transfer profile in the spatiotemporal dynamics can be characterized. As explained earlier, in this experiment, we drive the arm with the oscillatory motor command with an obstacle in the environment. We will see how the LMSIT ($LMSIT^{ST}_{S_i \to S_j}(\tau)$) captures this effect from the environment on the soft robotic arm. As a comparison, we will also drive the system without an obstacle in the environment with random motor commands and the same oscillatory motor command for 5000 timesteps and measure the LMSIT. In this experiment, we focus on the information transfer in the direction from the base to the tip (that is, $i < j$) by skipping the case when the time series of S_1 and S_{101} becomes the information sender or receiver under consideration, taking into account the results of experiment 1. In addition, similar to the setting of MSIT in experiment 1, N_R and T_R are both set to 1. If the information sender S_i is in the neighboring point of S_j, we need to adjust the corresponding $\hat{v}^{ST}_{i,j,t-\tau}$. Again, this procedure is described in the Appendix.

We start by observing the result for $LMSIT^{ST}_{S_i \to S_j}(\tau)$ when $i = j - 1$ and $\tau = 1$. Results are shown in the upper two plots of each row in Fig. 7 (a). The first upper line shows the example of the time series in the case of the random motor command, the oscillatory motor command, and the same oscillatory motor command with the obstacle, from left to right. The second upper line shows the corresponding $LMSIT^{ST}_{S_i \to S_j}(\tau)$ with

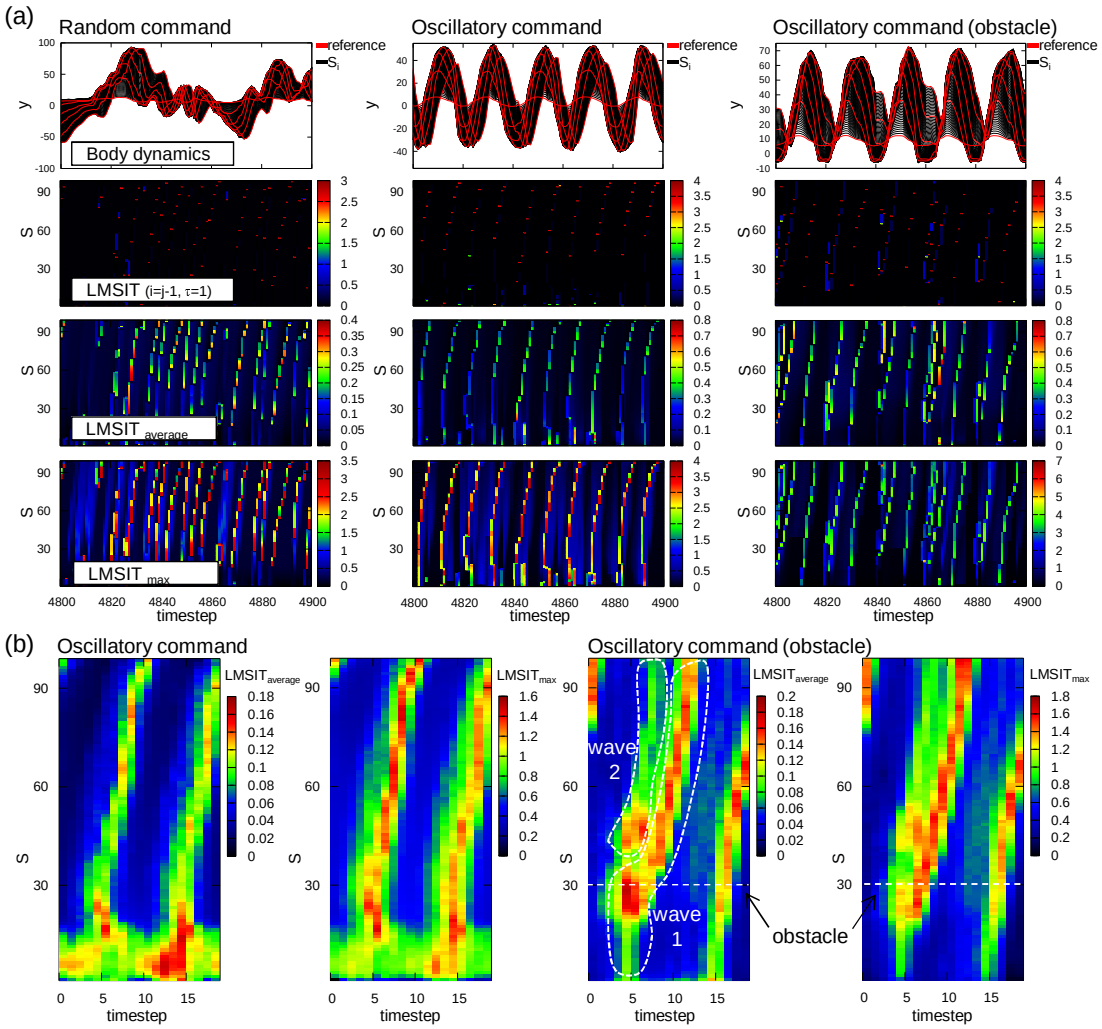


Figure 7. Results for the LMSIT analyses. (a) Typical LMSIT profiles corresponding to the body dynamics for the case of the random motor command (left row), the oscillatory motor command (middle row), and the oscillatory motor command with an obstacle (right row). For each row, the overlaid plot for the S_i time series, the corresponding LMSIT profile with $i = j - 1$ and $\tau = 1$, $LMSIT_{average}$ profile, and $LMSIT_{max}$ are presented from the top to the bottom. (b) The average $LMSIT_{average}$ and $LMSIT_{max}$ profile for the case with the oscillatory motor command without (left side) and with the obstacle (right side). Note that the timesteps when the motor commands switch to drive the arm toward the lower ($y < 0$) and the upper ($y > 0$) region in the screen correspond to timestep 5 and timestep 15, respectively, in both cases. The white line in the case with the obstacle represents the position of the contact point in relation to the obstacle. The white closures in the case with the obstacle depicts two spatially split information propagation waves, which we call *wave 1* and *wave 2*, respectively. See the main text for details.

$i = j - 1$ and $\tau = 1$. In each plot, we observe high $LMSIT^{ST}_{S_i \to S_j}(\tau)$ instantaneously in several spatiotemporal points. However, the pattern of the information structure is not obvious. To visualize the information transfer profile more clearly, by varying the delay τ from 1 to 20 and the information sender S_i for all the body parts in $i < j$, we measured the average LMSIT ($LMSIT_{average}(j) = \frac{1}{20}\sum_{\tau=1}^{20}\frac{1}{j-1}\sum_{d=1,j-d<j}^{j-1} LMSIT^{ST}_{S_{j-d}\to S_j}(\tau)$) and the maximum LMSIT ($LMSIT_{max}(j) = \max_{1\le\tau\le 20, 1\le d\le j-1}[LMSIT^{ST}_{S_{j-d}\to S_j}(\tau)]$) for each information receiver S_j. The results for $LMSIT_{average}(j)$ and $LMSIT_{max}(j)$ are shown in the lower two plots in Fig. 7 (a). In the case for the random motor command, we can clearly observe a high information propagation wave from the base toward the tip at the irregular temporal point. In the case of the oscillatory motor command, we observe a similar high information propagation wave from the base toward the tip two times in each oscillatory cycle at the specific timing. In the case of the oscillatory motor command with the obstacle, we observe a clear high information propagation wave from the base toward the tip similar to the case with the oscillatory motor command without the obstacle in half of each oscillatory cycle, but when the arm hits the obstacle in the other half of each oscillatory cycle, we see the fluctuating information propagation wave, where sometimes the wave splits into two, from the base toward the tip.

To see this fluctuating information propagation wave in detail, we averaged the $LMSIT_{average}(j)$ and $LMSIT_{max}(j)$ for each spatiotemporal point using 150 oscillation cycles for the oscillatory motor command with and without an obstacle (Fig. 7 (b)). As is obvious from the plots, in the case without the obstacle, we can clearly confirm two strong information propagation waves from the base toward the tip at specific timing, as we saw in Fig. 7 (a). In the case with the obstacle, we see that at the onset of the obstacle crush (around timestep 5), two spatially split strong information transfers are detected around the contact point of the object with the arm, and this information propagates toward the tip. We speculate that one propagation wave from the base to the tip (wave 1) represents the propagation of the motor actuation through the body, and the other wave (wave 2) represents the fast shock wave due to the contact to the obstacle. Furthermore, we can see that the velocity of wave 1 is modulated at the contact point (Fig. 7 (b)). This implies that the environmental damage provided by the crush to the obstacle induces a qualitatively different information transfer profile through the body from the one generated by the motor actuation. We expect that LMSIT is capable of detecting the shock wave for the case of multiple-point contact as long as its effect is reflected in the body dynamics.

4. Discussion

In this paper, by using the physical soft robotic arm platform, we demonstrated that the information-theoretic approach can be effective in characterizing the diverse spatiotemporal dynamics of soft bodies especially with delayed interaction. Currently, TE is one of the most applied measures to assess information transfers in robotics. However, it has drawbacks in capturing delayed interaction and applying to multivariate

time series. In fact, these points are the major properties of soft body dynamics, which have high-dimensionality and memory. To overcome these weak points, in this paper, we have introduced measures MSIT and LMSIT for applications to spatiotemporal dynamics of soft robots. By using MSIT, we showed that the actuation to the arm transfers the information toward the tip with a specific time delay, and we evaluated this information transfer velocity. Furthermore, we captured the delayed information transfer between each body part in detail. By using LMSIT, we visualized the dynamic information transfer profile hidden in the spatiotemporal dynamics of the soft body, and characterized the environmental damage that spreads throughout the entire body. These measures and the scheme can be applied effectively to soft robotic platforms in general.

Although we illustrated one possible scenario to characterize the information structure in a soft robotic platform, our approach can be extended in several ways taking into account each soft robotic platform in use and each experimental condition. For example, in our experiment, we set the relation between space and time based on the sampling rate of the experimental devices. However, the ratio between the experimentally observed space (the points in the body) and time does not always match the underlying physical dynamics, and this may make our proposed scheme ineffective. In those situations, experimenters should control the spatiotemporal scale of the experimentally obtained data. In addition, as we saw in the definition of M(S)IT, that one important point to capture the delayed information transfer was to condition on other causal contributors. Although we have introduced a reasonable physical assumption for conditioning, considering that soft robots can take unconventional morphology, it would be valuable in some situations to apply a method for reconstructing all the causal contributors from scratch only by using the obtained time series, which is introduced in [25, 26]. On the use of MIT, it has been recently reported in [31] that there is a case in which MIT does not effectively work in capturing the delayed couplings. In [31], an alternative measure called TE_{SPO}, where SPO stands for *self-prediction optimality*, has been proposed to measure the delayed information transfer. Their approach can be also applied for the analyses. Nevertheless, their measure is defined for the bivariate case. To extend the measure for the multivariate case, the concept presented in [25, 26] would be definitely useful and inevitable.

5. Conclusion

The behavior of animals or robots is generated by the dynamic coupling between the brain (controller), the body, and the environment. These dynamic interactions between components are usually described based on separately defined dynamical systems (such as the neural network for the brain and physics and mechanics for the body). When illustrating a consistent view of the dynamic interaction modality among these components, several difficulties will arise, for example, in determining the time scales among the components and the threshold for modeling the physical contact. In

this case, the information-theoretic approach can provide an effective method for giving a consistent view to characterize these dynamic couplings as information transfers, since the scheme is intrinsically based on a model-free approach. In particular, it has been recently shown that soft bodies can be exploited as a computational resource (see, e.g., [20, 21]). In this case, the soft body itself acts similarly to a huge recurrent neural network. Although the brain has the apparent anatomical structure (network topology) to realize interactions among neurons, the body does not contain this type of explicit interaction pathway but has a characteristic morphology and can physically interact with the environment. Our approach presented here would be also useful to reveal how each body part interacts with another, which is usually hidden in the dynamics, and to reconstruct a functional network topology within the body. This type of analysis would be valuable to infer the information processing capacity of the body and is included in our future work. Moreover, it would be valuable to investigate how the body morphology and the controller of soft robots self-organize in a given environment, guided by information transfers [60]. For example, by focusing on several variables in a soft robotic platform and by maximizing the information transfer between these variables, we can investigate how the design of the robot and the controller coevolves [61], which would reveal the relations between the dynamic coupling and the resulting behavioral generation and help understanding a design principle for soft robots.

Acknowledgments

KN would like to acknowledge Dr. Tao Li for his help in collecting the experimental data and Dr. Hidenobu Sumioka for fruitful discussions in the early stage of the study. KN, RF was supported by the European Commission in the ICT-FET OCTOPUS Integrating Project (EU project FP7-231608). KN was also supported by JSPS Postdoctoral Fellowships for Research Abroad. NS, RF was supported by EU project Extending Sensorimotor Contingencies to Cognition (eSMCs), IST-270212.

Appendix

Time series data preparation

In this Appendix, we provide details on the time series preparation for S. As explained in the main text, we first track the positions of 6 reference points in the body, (x_{Ri}, y_{Ri}) $(i = 1, 2, ..., 6)$, for each timestep (Fig. 3 (a)) and fitted these points with a 5th order polynomial curve, expressed as

$$y = a_5 x^5 + a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0, \tag{20}$$

where a_5, a_4, a_3, a_2, a_1, and a_0 are the constant parameters to be tuned (Fig. 3 (b)). The choice of the order of a polynomial curve was determined by observing the arm motion, where the number of bends does not exceed what can be expressed as a 5th order polynomial curve. Let us introduce a vector expression for these parameters,

$\mathbf{A} = [a_5, a_4, a_3, a_2, a_1, a_0]^{\mathrm{T}}$. By using the x-coordinates of the collected reference points, a matrix $\mathbf{X}$ is denoted as

$$\mathbf{X} = \begin{bmatrix} x_{R1}^5 & x_{R1}^4 & x_{R1}^3 & x_{R1}^2 & x_{R1} & 1 \\ x_{R2}^5 & x_{R2}^4 & x_{R2}^3 & x_{R2}^2 & x_{R2} & 1 \\ x_{R3}^5 & x_{R3}^4 & x_{R3}^3 & x_{R3}^2 & x_{R3} & 1 \\ x_{R4}^5 & x_{R4}^4 & x_{R4}^3 & x_{R4}^2 & x_{R4} & 1 \\ x_{R5}^5 & x_{R5}^4 & x_{R5}^3 & x_{R5}^2 & x_{R5} & 1 \\ x_{R6}^5 & x_{R6}^4 & x_{R6}^3 & x_{R6}^2 & x_{R6} & 1 \end{bmatrix}.$$

Then, denoting the y-coordinates of the collected reference points as $\mathbf{Y} = [y_{R1}, y_{R2}, y_{R3}, y_{R4}, y_{R5}, y_{R6}]^{\mathrm{T}}$, we can approximate the constant parameters using the Moore-Penrose pseudo-inverse as $\mathbf{A} = \mathbf{X}^{-1}\mathbf{Y}$. Using the fitted curve, we fragmented the interval of each neighboring reference point into 20 equidistant segments along the curve (Fig. 3 (c)). The resulting 101 endpoints for these segments are denoted from S_1 to S_{101} from the base to the tip, including the reference points. By iterating these procedures for each timestep, we collect y-coordinates for these endpoints resulting in 101 time series for $S_1, ..., S_{101}$ (Fig. 3 (c)). These time series are the representation of the body dynamics and are used for the analyses.

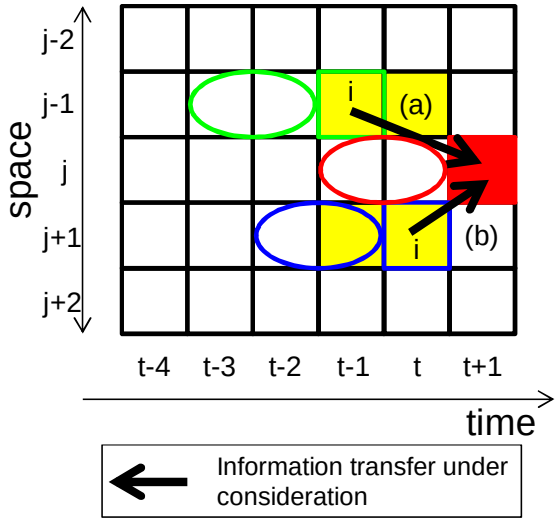


Figure 8. Schematics explaining the setting of $\hat{v}^{ST}_{i,j,t-\tau}$ when the information sender i is overlapped with $\hat{v}^{ST}_{i,j,t-\tau}$ through embedding dimensions. Here, $\hat{v}^{ST}_{i,j,t-\tau}$ is usually set as $\hat{v}^{ST}_{i,j,t-\tau} = \{\hat{x}^2_{j+1,t}, \hat{x}^2_{j-1,t}\}$ from the definition. (a) Schematic expressing information transfer from the sender $i = j-1$ to the receiver j with delay $\tau = 2$. In this case, $\hat{v}^{ST}_{i,j,t-\tau}$ becomes to $\hat{v}^{ST}_{i,j,t-\tau} = \{\hat{x}^2_{j+1,t}\}$. (b) Schematic expressing information transfer from the sender $i = j+1$ to the receiver j with delay $\tau = 1$. In this case, $\hat{v}^{ST}_{i,j,t-\tau}$ becomes to $\hat{v}^{ST}_{i,j,t-\tau} = \{\hat{x}^2_{j-1,t}\}$. For each information sender and receiver, the circle expresses the condition under its own past. Note that, in the figure, $(K, L, M, N_r, T_r) = (2, 3, 2, 1, 1)$.

Settings of possible information sources in our analyses

Here, we explain how we set the condition of possible information sources $\hat{v}^{ST}_{i,j,t-\tau}$ ($\hat{v}^{ST}_{i,j,t}$) in our analyses. For descriptive purposes, we unify the expression of possible information sources as $\hat{v}^{ST}_{i,j,t-\tau}$ in this section. In our analyses, if the information sender has an overlap with $\hat{v}^{ST}_{i,j,t-\tau}$ through embedding dimensions, then we discarded the term containing the overlap in $\hat{v}^{ST}_{i,j,t-\tau}$. Typical cases for this setting are exemplified in Fig. 8.

References

[1] Pfeifer R, Lungarella M and Iida F 2012 The challenges ahead for bio-inspired 'soft' robotics *Commun. ACM.* **55** 76-87

[2] Trivedi D, Rahn CD, Kier WM and Walker ID 2008 Soft robotics: biological inspiration, state of the art, and future research *Appl. Bionics. Biomech.* **5** 99-117

[3] Kim S, Laschi C and Trimmer B 2013 Soft robotics: a new perspective in robot evolution *Trends Biotechnol.* **31** 287-94

[4] Shepherd RF, Ilievski F, Choi W, Morin SA, Stokes AA, Mazzeo AD, Chen X, Wang M and Whitesides GM 2011 Multi-gait soft robot *Proc. Natl. Acad. Sci. U. S. A.* **108** 20400-3

[5] Pfeifer R, Lungarella M and Iida F 2007 Self-organization, embodiment, and biologically inspired robotics *Science* **318** 1088-93

[6] Pfeifer R and Bongard J 2006 *How the Body Shapes the Way We Think: A New View of Intelligence* (MIT Press, Cambridge, MA)

[7] Nakajima K, Ngouabeu AMT, Miyashita S, Göldi M, Füchslin RM and Pfeifer R 2012 Morphology-Induced Collective Behaviors: Dynamic Pattern Formation in Water-Floating Elements *PLoS ONE* **7** e37805

[8] Hochner B 2012 An embodied view of octopus neurobiology *Curr. Biol.* **22** R887-92

[9] Sumbre G, Fiorito G, Flash T and Hochner B 2005 Motor control of flexible octopus arms *Nature* **433** 595-6

[10] Sumbre G, Gutfreund Y, Fiorito G, Flash T and Hochner B 2001 Control of octopus arm extension by a peripheral motor program *Science* **293** 1845-8

[11] Laschi C, Mazzolai B, Mattoli V, Cianchetti M and Dario P 2009 Design of a biomimetic robotic octopus arm *Bioinsp. Biomim.* **4** 015006

[12] Cianchetti M, Arienti A, Follador M, Mazzolai B, Dario P and Laschi C 2011 Design concept and validation of a robotic arm inspired by the octopus *Mater. Sci. Eng. C* **31** 1230-9

[13] Laschi C, Mazzolai B, Cianchetti M, Margheri L, Follador M and Dario P 2012 A soft robot arm inspired by the octopus *Adv. Robot.* **26** 709-27

[14] Calisti M, Giorelli M, Levy G, Mazzolai B, Hochner B, Laschi C and Dario P 2011 An octopus-bioinspired solution to movement and manipulation for soft robots *Bioinsp. Biomim.* **6** 036002

[15] Li T, Nakajima K, Calisti M, Laschi C and Pfeifer R 2012 Octopus-Inspired Sensorimotor Control of a Multi-Arm Soft Robot *Proc. Int. Conf. on Mechatronics and Automation* (Chengdu, China, 5-8 August) pp 948-55

[16] Li T, Nakajima K, Kuba MJ, Gutnick T, Hochner B and Pfeifer R 2011 From the octopus to soft robots control: an octopus inspired behavior control architecture for soft robots *Vie et Milieu* **61** 211-7

[17] Kuwabara J, Nakajima K, Kang R, Branson DT, Guglielmino E, Caldwell DG and Pfeifer R 2012 Timing-Based Control via Echo State Network for Soft Robotic Arm *Proc. Int. Joint Conf. on Neural Networks* (Brisbane, Australia, 10-15 June) pp 1-8

[18] Li T, Nakajima K and Pfeifer R 2013 Online Learning Technique for Behavior Switching in a Soft Robotic Arm *Proc. IEEE Int. Conf. on Robotics and Automation* (Karlsruhe, Germany, 6-10 May) pp 1296-1302

[19] Li T, Nakajima K, Cianchetti M, Laschi C and Pfeifer R 2012 Behavior Switching by Using Reservoir Computing for a Soft Robotic Arm *Proc. IEEE Int. Con. on Robotics and Automation* (Saint Paul, MN, USA, 14-18 May) pp 4918-24

[20] Nakajima K, Hauser H, Kang R, Guglielmino E, Caldwell DG and Pfeifer R 2013 A soft body as a reservoir: case studies in a dynamic model of octopus-inspired soft robotic arm *Front. Comput. Neurosci.* **7** 1-19

[21] Nakajima K, Hauser H, Kang R, Guglielmino E, Caldwell DG and Pfeifer R 2013 Computing with a Muscular-Hydrostat System *Proc. IEEE Int. Conf. on Robotics and Automation* (Karlsruhe, Germany, 6-10 May) pp 1504-11

[22] Amblard PO and Michel OJJ 2011 On directed information theory and Granger causality graphs *J. Comput. Neurosci.* **30** 7-16

[23] Marschinski R and Kantz H 2002 Analysing the information flow between financial time series - an improved estimator for transfer entropy *Eur. Phys. J. B* **30** 275-281

[24] Pompe B and Runge J 2011 Momentary information transfer as a coupling measure of time series *Phys. Rev. E* **83** e051122

[25] Runge J, Heitzig J, Petoukhov V and Kurths J 2012 Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy *Phys. Rev. Lett.* **108** 258701

[26] Runge J, Heitzig J, Marwan N and Kurths J 2012 Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy *Phys. Rev. E* **86** 061121

[27] Schreiber T 2000 Measuring information transfer *Phys. Rev. Lett.* **85** 461-4

[28] Schmidt N, Hoffmann M, Nakajima K and Pfeifer R 2013 Bootstrapping Perception Using Information Theory: Case Studies in a Quadruped Robot Running on Different Grounds *Adv. Complex Syst.* **16** 1250078

[29] Lungarella M and Sporns O 2006 Mapping information flow in sensorimotor networks *PLoS Comput. Biol.* **2** 1301-12

[30] Cover TM and Thomas JA 1991 *Elements of Information Theory* (Wiley-Interscience)

[31] Wibral M, Pampu N, Priesemann V, Siebenhühner F, Seiwert H, Lindner M, Lizier JT and Vicente R 2013 Measuring Information-Transfer Delays *PLoS ONE* **8** e55809

[32] Ay N and Polani D 2008 Information Flows in Causal Networks *Adv. Complex Syst.* **11** 17-41

[33] Lizier JT and Prokopenko M 2010 Differentiating information transfer and causal effect *Eur. Phys. J. B* **73** 605-615

[34] Bandt C and Pompe B 2002 Permutation entropy: A natural complexity measure for time series *Phys. Rev. Lett.* **88** e174102

[35] Staniek M and Lehnertz K 2008 Symbolic transfer entropy *Phys. Rev. Lett.* **100** e158101.

[36] Bahraminasab A, Ghasemi F, Stefanovska A, McClintock PVE and Kantz H 2008 Direction of coupling from phases of interacting oscillators: A permutation information approach *Phys. Rev. Lett.* **100** e084101

[37] Rosso OA, Larrondo HA, Martin MT, Plastino A and Fuentes MA 2007 Distinguishing noise from chaos *Phys. Rev. Lett.* **99** e154102

[38] Kugiumtzis D 2013 Partial transfer entropy on rank vectors *Eur. Phys. J. Special Topics* **222** 401-20.

[39] Nakajima K and Haruna T 2013 Symbolic local information transfer *Eur. Phys. J. Special Topics* **222** 421-39

[40] Cao YH, Tung WW, Gao JB, Protopopescu VA and Hively LM 2004 Detecting dynamical changes in time series using the permutation entropy *Phys. Rev. E* **70** e046217

[41] Bandt C, Keller G and Pompe B 2002 Entropy of interval maps via permutations *Nonlinearity* **15** 1595-1602

[42] Amigó JM 2010 *Permutation Complexity in Dynamical Systems* (Springer-Verlag: Berlin, Heidelberg, Germany)

[43] Amigó JM, Kennel MB, Kocarev L 2005 The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems *Physica D* **210** 77-95

[44] Amigó JM, Keller K 2013 Permutation entropy: One concept, two approaches *Eur. Phys. J. Special Topics* **222** 263-73
[45] Haruna T and Nakajima K 2011 Permutation complexity via duality between values and orderings *Physica D* **240** 1370-7
[46] Amigó JM 2012 The equality of Kolmogorov-Sinai entropy and metric permutation entropy generalized *Physica D* **241** 789-93
[47] Keller K, Unakafov AM and Unakafova VA 2012 On the relation of KS entropy and permutation entropy *Physica D* **241** 1477-81
[48] Unakafova VA, Unakafov AM and Keller K 2013 An approach to comparing Kolmogorov-Sinai and permutation entropy *Eur. Phys. J. Special Topics* **222** 353-61
[49] Keller K and Sinn M 2009 A standardized approach to the Kolmogorov-Sinai entropy *Nonlinearity* **22** 2417-22
[50] Keller K and Sinn M 2010 Kolmogorov-Sinai entropy from the ordinal viewpoint *Physica D* **239** 997-1000
[51] Keller K 2012 Permutations and the Kolmogorov-Sinai entropy *Discr. Cont. Dyn. Syst.* **32** 891-900
[52] Haruna T and Nakajima K 2013 Permutation Complexity and Coupling Measures in Hidden Markov Models *Entropy* **15** 3910-30
[53] Haruna T and Nakajima K 2013 Permutation approach to finite-alphabet stationary stochastic processes based on the duality between values and orderings *Eur. Phys. J. Special Topics* **222** 367-83
[54] Haruna T and Nakajima K 2013 Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic Markov processes *Eur. Phys. J. B* **86** 230
[55] Kugiumtzis D 2012 Transfer entropy on rank vectors *J. Nonlin. Sys. Appl.* **3** 73-81.
[56] Lizier JT, Prokopenko M, Zomaya AY 2008 Local information transfer as a spatiotemporal filter for complex systems *Phys. Rev. E* **77** 026110
[57] Nakajima K, Li T, Kang R, Guglielmino E, Caldwell DG and Pfeifer R 2012 Local Information Transfer in Soft Robotic Arm *Proc. IEEE Int. Conf. on Robotics and Biomimetics* (Guangzhou, China, 11-14 December) pp 1273-80
[58] Brown D, Tracker video analysis and modeling tool. http://www.cabrillo.edu/ dbrown/tracker/, December, 2009.
[59] Nakajima K, Li T, Sumioka H, Cianchetti M and Pfeifer R 2011 Information Theoretic Analysis on a Soft Robotic Arm Inspired by the Octopus *Proc. IEEE Int. Con. on Robotics and Biomimetics* (Phuket, Thailand, 7-11 December) pp 110-7
[60] Martius G, Der R and Ay N 2013 Information driven self-organization of complex robotic behaviors. *PLoS ONE* **8** e63400
[61] Bongard J 2011 Morphological change in machines accelerates the evolution of robust behavior. *Proc. Natl. Acad. Sci. U. S. A.* **108** 1234-9

Appendix D

Using Sensorimotor Contingencies for Terrain Discrimination and Adaptive Walking Behavior in the Quadruped Robot Puppy

Reprinted from:

Hoffmann, M., Schmidt, N. M., Pfeifer, R., Engel, A. K. and Maye, A. (2012). "Using sensorimotor contingencies for terrain discrimination and adaptive walking in the quadruped robot Puppy." In Ziemke, T., Balkenius, C., and Hallam, J. (Eds.), *From animals to animats 12: Proc. Int. Conf. Simulation of Adaptive Behavior (SAB)*, Odense, Denmark, Vol. 7246 of LNAI, Springer, pages 54–64. doi:10.1007/978-3-642-33093-3_6.

This is the final accepted version. The original publication is available at www.springerlink.com (http://link.springer.com/chapter/10.1007/978-3-642-33093-3_6).

Using Sensorimotor Contingencies for Terrain Discrimination and Adaptive Walking Behavior in the Quadruped Robot Puppy

Matej Hoffmann[1], Nico M. Schmidt[1], Rolf Pfeifer[1], Andreas K. Engel[2], and Alexander Maye[2]

[1] Artificial Intelligence Laboratory, Department of Informatics, University of Zurich
Andreasstrasse 15, 8050 Zurich, Switzerland
{hoffmann,nschmidt,pfeifer}@ifi.uzh.ch
[2] University Medical Center Hamburg-Eppendorf
Dept. of Neurophysiology and Pathophysiology
Martinistr. 52, 20246 Hamburg, Germany
{a.maye,ak.engel}@uke.de

Abstract. In conventional "sense-think-act" control architectures, perception is reduced to a passive collection of sensory information, followed by a mapping onto a prestructured internal world model. For biological agents, Sensorimotor Contingency Theory (SMCT) posits that perception is not an isolated processing step, but is constituted by knowing and exercising the law-like relations between actions and resulting changes in sensory stimulation. We present a computational model of SMCT for controlling the behavior of a quadruped robot running on different terrains. Our experimental study demonstrates that: (i) Sensory-Motor Contingencies (SMC) provide better discrimination capabilities of environmental properties than conventional recognition from the sensory signals alone; (ii) discrimination is further improved by considering the action context on a longer time scale; (iii) the robot can utilize this knowledge to adapt its behavior for maximizing its stability.

Keywords: active perception, terrain recognition, object recognition, developmental robotics, adaptive behavior

1 Introduction

In the majority of approaches to robot control the extraction and classification of features from the sensory input is a crucial processing step that has a critical effect on the behavioral performance of the artificial agent. Ever more complex methods are employed to detect type and position of objects, to recognize landmarks and obstacles, or to infer the spatial configuration of the surrounding area. In mobile robotics, for example, this problem is typically solved by employing several distal (non-contact) sensors: cameras, laser range finders, and possibly also radar. Terrain classification into traversable vs. non-traversable is done in

a supervised manner through a set of labeled terrain examples [1]. This is used to update an internal representation of the world – a 2D occupancy grid that in turn is used for planning a collision-free path. Although recent studies suggest that the traditional "sense-think-act" approaches can also be extended to real-world environments, their task domain is still limited.

The inherent problem of these approaches, in our view, is that they treat perception as a separate, *passive* process that is detached from the agent's actions. A "sensory snapshot" of the environment is taken that is then mapped onto the states of an internal world model. However, we believe that perception in biological agents has a different character. First, it is active. This view can be traced back to the pragmatic philosopher John Dewey [3], and it was later picked up by research in active perception (see [4] for an overview). Second, perception occurs through the body. The information that reaches the brain is thus critically shaped by the active generation of sensory stimuli and by the agent's embodiment (this is quantified in [10], for instance). Sensorimotor Contingency Theory (SMCT)[15, 14] as a representative of action-oriented approaches ascribes sensory awareness and perception to the exercise of knowledge about the lawful relations between actions and resulting changes in the sensory signals, called Sensory-Motor Contingencies (SMCs), instead of activating an internal representation of the perceived object.

We have recently developed a computational model for SMCs and demonstrated its application in an object-recognition task [11]. Here we apply the same model for controlling a robot with a completely different embodiment: a quadruped "dog" robot. We start by investigating how different gaits and terrains modulate the sensory information collected by the robot. Next we demonstrate that taking the action explicitly into account improves the terrain classification accuracy. Taking the context of longer sensorimotor sequences into account can further improve the classification performance. Finally, we show that the robot can successfully deploy its perception of the properties of different grounds to select gaits from a given repertoire to maximize its stability.

2 Related Work

The importance of sensorimotor information for object recognition in humans is evident from studies of neurological disorders [22], even though it is sometimes assigned only the role of a fall-back system [18]. In a scenario similar to ours, E.J. Gibson et al. [5] studied how infants perceive the traversability of the environment, implicitly taking into account their mode of locomotion – walking or crawling – and exploiting not only visual but also tactile information. In general, perceptual categorization in biological agents is a hard problem [7] resulting from a complex interplay of the brain, body and environment, and the individual effects are hard to separate. In this regard, robotics has provided efficient tools to test these effects independently.

First, Pfeifer and Scheier [16] have demonstrated how sensorimotor coordination can greatly simplify classification or categorization in a study where

mobile robots distinguish between big and small cylinders by circling around them. Whereas this would be very difficult from a static camera picture when the distance to the object is not known, different angular velocities resulting from circling around them render the problem much easier. Similar results emerged from studies in artificial evolution: the fittest agents were those engaging in sensory-motor coordinated behavior [2].

Second, perception can be facilitated by the morphology of the body and the sensory apparatus (see examples in [8]). In legged robots that engage in different terrains, proprioceptive sensors can be particularly useful. In a previous study in our platform, we have shown how information regarding the robot's position and orientation can be extracted [17]. A combination of proprioceptive sensors has been successfully employed in a terrain recognition task in a hexapod [6].

Third, the action that caused a sensory stimulation can be explicitly taken into account in a classification task. This has been done in [19], where sensory data resulting from different actions are clustered separately. In [20], traversability categories are predefined and the robot learns – for each action separately – a mapping from initial percepts to these categories.

Many more approaches employ some form of sensorimotor information, but to our knowledge the approach we will present here is one of the few in that actions play a constitutive role for the perception of the agent as proposed by SMCT. Our method allows for a context given by the sequence of previous actions, and it is inherently multimodal. In addition, we will test the hypothesis that longer sensorimotor sequences are needed for object categorization (i.e., the ground the robot is running on in our case). Furthermore, to demonstrate the behavioral relevance of the classification capabilities for the agent, we present a closed-loop system that employs the perception of the properties of different grounds to select gaits from a given repertoire to maximize stability.

3 Methods and experiments

3.1 Robot and Experimental Setup

The Puppy robot (see Fig. 1 left) has four identical legs driven by position-controlled servomotors in the hips. It has passive compliant joints at the knees. We prepared five sets of position control commands for the servomotors, resulting in five distinct gaits (bound forwards, bound left/right, crawl, trot backwards), each of them with a periodic motor signal at 1 Hz. Four potentiometers measured the joint angles on the passive knee joints, and 4 pressure sensors recorded forces applied to the robot's feet. Linear accelerations (in X, Y, and Z direction) were measured by an onboard accelerometer. In total we used 11 sensory channels, jointly sampled at 50Hz.

To investigate the long-term properties of our approach, we additionally designed a model of Puppy in Webots [21], a physics-based simulator (see Fig. 1 right). For this model we used the same gait repertoire (2 gaits had to be

adapted) plus 4 additional gaits (turn left/right, pace, walk), obtaining a repertoire of nine gaits. In both cases, gaits (actions) were exercised in 2-second-intervals during which the sensory data were collected, forming sensorimotor epochs of 2 seconds. At the end of each epoch the robot could change the gait.

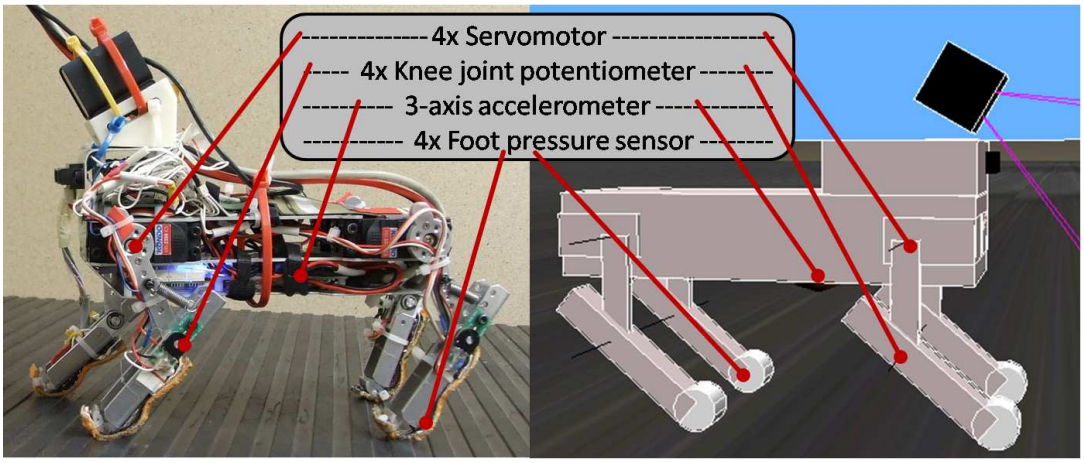


Fig. 1. Real and simulated robot and the sensor suite. The robot is 20 cm long. The camera and infrared sensors that are also mounted on the robot were not used in the experiments.

For the real robot, we prepared a small wall-enclosed arena of 2x1 m. Four different ground substrates covered the ground: plastic foil, cardboard, Styrofoam and rubber. These materials differed in friction and also in structure (cardboard and rubber had ridges). In the simulator, the arena was much bigger in size (25x25 m), so encounters with the walls were much less frequent. The "foil", "cardboard", and "rubber" were flat but differed in Coulomb friction coefficients (μ = 2, 11, and 20 respectively). To increase the differences between the substrates in the simulator, the "Styrofoam" ground (μ = 9) was made uneven with randomly placed smooth bumps of up to 3 cm height.

3.2 Feature Computation

For effective processing of sensorimotor information we compressed the raw data by extracting some simple features. For the action space we chose a high abstraction level and used the gait as a single feature. In the sensory space, following a similar strategy as we used in [17], we took advantage of the periodic nature of the locomotion and created period-based features as follows: (1) sum of knee amplitudes of all legs in a period,[3] (2) sum of standard deviations of all knee joints, (3) sum of mean pressures in each foot, (4) sum of standard deviations of each foot pressure signal, (5-7) mean accelerations along X,Y, and Z-axis respectively, (8-10) standard deviations of the accelerometer signals. Since frequent

[3] Note that the knees are passive compliant.

gait transitions disrupt the locomotion and impact also the sensory values, only the last second (i.e. the second locomotion period) from each 2s epoch was used for the feature computation. Continuous feature values were used for classification (Section 4.1); for learning the sensorimotor contingencies and optimizing the behavior using a Markov model (Section 4.2), each feature was quantized to two levels only.

3.3 A Markov Model of SMCs

We employed the model that we presented in [11, 12] with the necessary adaptations to the Puppy robot. The basic idea is to consider actions and resulting changes in sensory signals in an integrated manner, and to keep a record of sequences of actions and sensory observations. For each epoch, the action a (the gait in this case) and a vector of n sensory features observed during execution of a are concatenated to a single vector $ao(t) = [as_1s_2 \dots s_n]$ that we call an action-observation pair. Based on the sequence of action-observation tuples that the robot experiences over time $c^h = [ao(t), ao(t-1), \dots ao(t-h)]$, the model samples the conditional probability distributions $P^h(ao(t+1)|c^h(t))$, i.e. the probability of experiencing a particular action-observation pair in the next time step given a finite history h of previous pairs. In this study we use $h = 0 \dots 4$. This probability distribution is what we call the extended Sensori-Motor Contingencies (eSMC) of an agent, and a particular combination of $ao(t+1)$ and $c^h(t)$ is a specific sample that in addition to its probability of occurrence can have other properties like a value.

3.4 Value System and Action Selection

We extended the basic idea of SMCT by a value system and an action selection algorithm. For each epoch t, we define the value[4] of the robot's state by a weighted sum of three components:

$$v(t) = -tumbled - 0.4regularity - 0.1speed$$

We used the signal of the accelerometer in Z direction to determine if the robot is upright ($tumbled = 0$) or has tipped over ($tumbled = 1$). The similarity of the sensory patterns at the knee joints between the first and second period during an epoch is reflected in the *regularity* value (1 for identical patterns during both periods), and the normalized velocity computed from the robot's global coordinates yields the *speed* value.

We have devised a stochastic action selection algorithm that attempts to optimize the temporal average of the internal value. It selects actions that have shown to activate eSMCs with high internal values, and explores the consequences of new actions when no or only bad prior experiences exist in a given

[4] In reinforcement learning terms, this would be called reward - it is the immediate reward signal associated with each state.

situation. For each action-observation sequence $c^h(t)$ a record of actions executed next $a_{next}(c^h(t))$ and the average value $v(a_{next}(c^h(t)) = \sum_n v(t+1)/n$ is kept, where n is the number that action a_{next} was executed when context $c^h(t)$ was encountered, and $v(t+1)$ is the resulting value. Different history lengths h may yield different value information. Since we consider longer matches between a particular action-observation sequence and the stored eSMCs as a more accurate estimation of the state, preference is given to the value information from longer matching histories. When the robot later experiences the same context again, it knows the average value of the actions it has tried before. Random values get assigned to the other actions. To avoid a predominantly random exploration in the initial learning phase when the robot has only little sensorimotor knowledge, the expected value for the most recently executed action is given by the internal value of the last epoch. This favors the continuation of successful actions, and switching to another action otherwise. The action with the highest expected value $\hat{a} = \arg\max_a v(a_{next}(c^h(t))$ is then executed with a probability $p(\hat{a}) = v(\hat{a}) + 1$.

4 Results

4.1 Perception and Discrimination of Different Grounds

In this section, we want to quantitatively assess the effect of considering actions and the resulting changes in sensory stimulation in an integrated manner. First, we compare the respective influence of the action (the gait the robot is running with) and the environment on the sensory data. Second, focusing on the ground discrimination, we demonstrate how explicitly incorporating the action that has induced a sensory stimulation improves the environment classification. Finally, we study the effect of longer sensorimotor sequences, testing our hypothesis that these are required for object categorization, whereby, from the robot's perspective, different grounds correspond to different objects in our scenario.

We have collected data from the real (4 x 20 minutes, i.e. 4 x 600 epochs) and simulated version of the robot (4 x 4 hours, i.e. 4 x 7200 epochs) running separately on the different substrates. After every epoch a new action was chosen at random. If the robot tumbled, it was manually (real robot) or automatically (simulator) returned to an upright position at the same location and two epochs following this event were discarded. A reflex for backing up from the walls was built in. Epoches when the robot was backing up (frequent in the real robot) were not discarded but entered the regular learning process. A naïve Bayes classifier (diagonal covariance matrix estimate, stratified 10-fold cross-validation) was trained to classify either the action or the ground substrate given the sensory observations and actions during the previous epochs.

Ground and Gait Discrimination from Sensory Data Only. To assess the dependencies of the sensory signals from the gait or ground, respectively, we collapsed the data across gaits (for assessing ground effects) or across grounds

(for assessing gait effects). In the real Puppy, the classifier determined the correct gait from the sensory data in 72.4% of the cases, and in 81.6% in the simulation. In contrast, the ground recognition rates were lower, 67.2% for the real Puppy and 43.1% in the simulation (see also Fig. 2, top-most bars). This shows that gaits and grounds have a similarly strong effect on the sensory patterns in the real robot. In the simulation the different materials induce similar sensory patterns and hence, are difficult to distinguish. These figures serve as a baseline when we consider the classification of joint action and sensor information next.

Ground Discrimination Using Action Information. We separated the data into sets for each gait and classified the grounds on each set individually. Afterwards we averaged the ground recognition rate over all gaits. In comparison to the ground recognition using a single classifier, the action-dependent classification schema reaches an improved accuracy of 75.7% for the real robot. Considering only the gait yielding the best recognition rate, this value increases to 80.2%. In the simulation this increase is even more pronounced, from 43.1% to 62.9% and 78.3%, respectively (see Fig. 2, second bars from top). This indicates that taking the action that caused a sensory observation into account is more specific for the environmental condition than analyzing the sensory data alone.

Ground Discrimination Using Action Sequences. The sensorimotor patterns induced by a single action may often be similar even if the agent interacts with different objects. As suggested by SMCT, longer sequences of interaction with an object may be needed in order for the object to leave a unique "footprint". We confirmed this hypothesis by splitting the data further into sets for specific sequences of 2 or 3 consecutive actions, and averaging again over all sequences. The sensory feature vectors from consecutive epochs were concatenated. For a sequence of two gaits, the ground classification accuracy rises to 84.7% in the real robot, and to 70.6% in the simulation. Considering a sequence of 3 gaits further improves accuracy (see Fig. 2). Here, the gait sequence-specific classifier with the highest accuracy achieves a 100% recognition rate. This means that the sensorimotor patterns of this action sequence are apt for a reliable recognition of the different grounds.

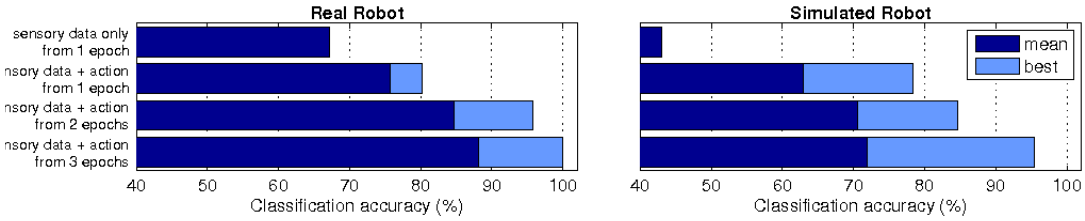


Fig. 2. Comparison of the ground classification accuracies when the action context is taken into account to different degrees. (left) Real robot. (right) Simulated robot.

4.2 Selecting Gaits to Optimize Behavior

Next we want to demonstrate how the better discrimination capabilities that result when a longer action context is considered can be used by the robot to improve its behavior. We let the simulated robot run on 4 different grounds, and used the Markov model (Sec. 3.3) to learn eSMCs for the 9 gaits from its repertoire. Each eSMC had an associated value given by the value function described in section 3.4.

With progressing sensorimotor knowledge, the robot preferred to choose gaits that improved its internal value, providing swift, smooth and stable locomotion. The plots of the value function in Fig. 3 show that a basic set of gaits that "feel good" to Puppy (i.e. maximize the value function) is found after only about 1.000 epochs (around 8 minutes). On cardboard it takes more than 2.000 epochs to arrive at a reasonable gait combination. Afterwards the robot tries to further improve its behavior by selecting from these comfortable gaits with different probabilities. As one would expect, the optimal gait sequence depends on the material properties of the grounds. Except for the plastic foil, Puppy prefers a mixture of walking back and turning left or right. It is most successful in epochs when it reduces the frequency of turns in favor of walking back. On plastic foil, the most successful gait is pacing, while turning left seems to be a less favorable gait. On cardboard, turning left is selected more frequently than turning right, though, while on rubber both turning actions are chosen with about the same frequency.

On the rough styrofoam, the value function is dominated by frequent tipping of the robot. Compared to the three flat grounds the value remains at a low level, and the separation into favorable and unpleasant gaits is less pronounced. The order of preference seems to be maintained, though.

The improvement of the internal value is not monotonic, but proceeds in a rather oscillatory manner. Intervals in which the robot had sufficient sensorimotor knowledge to optimize its behavior alternated with epochs in which it learned new eSMCs. With the sensorimotor knowledge growing, episodes with optimal behavior become more frequent and last longer. On cardboard, for example, behaviors that maximize the value function are found after about $2 \cdot 10^4$ epochs, and the increasing width of the peaks in the value function indicate that the robot spends more and more time in these optimal behaviors. A similar observation can be made on plastic foil. On rubber, the knowledge about favorable behavior around $2 \cdot 10^4$ seems to be lost afterwards, but it can be expected that the exploration process leads to a further improvement beyond the analyzed interval. Since the value function was designed to never reach zero, corresponding to a state of perfect harmony, the robot keeps on exploring the potential to further improve its fitness.

5 Conclusion and Future Work

In this study we have investigated sensorimotor classification of different substrates in a quadruped robot from the perspective of SMCT. First, we have

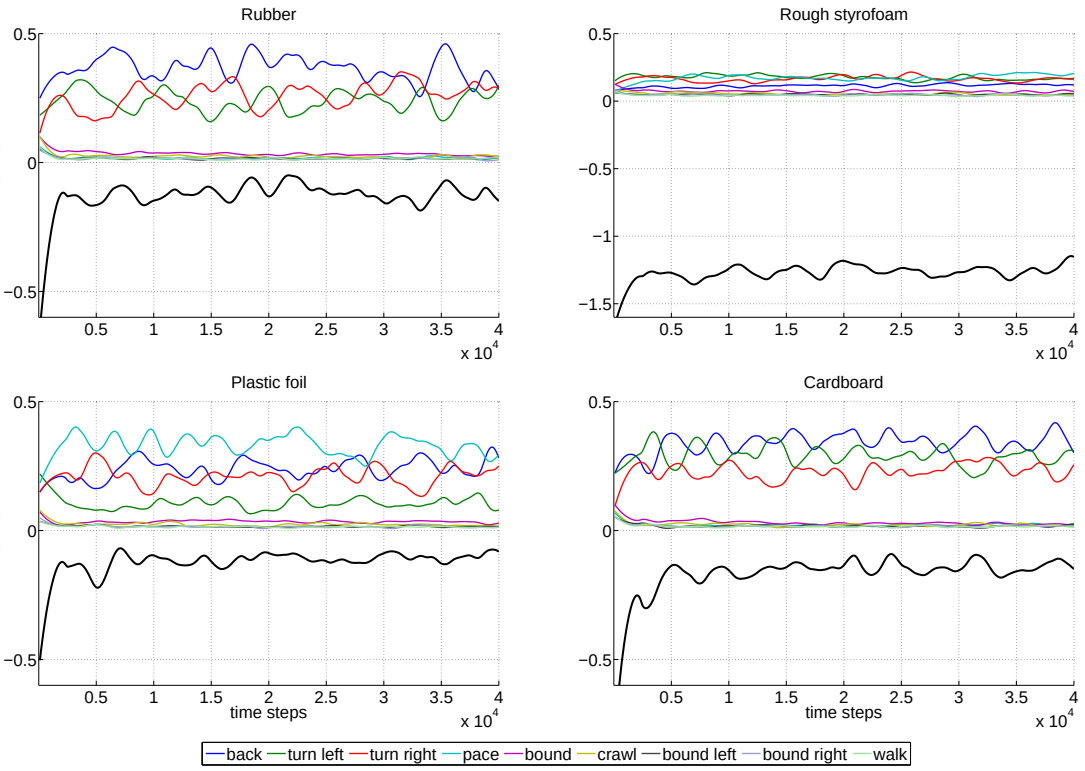


Fig. 3. Value (black curve under the abscissa) and gait selection frequencies (above) over time on 4 ground substrates (data from simulator). All curves have been smoothed with a weighted linear least squares and a 2nd degree polynomial model in a moving window of 5.000 samples. To appreciate the time course of the value function, the initially low values have been clipped. Note the different scale of the value function for rough styrofoam.

demonstrated how sensory stimulation patterns critically depend on the actions the robot is exercising. If the robot wants to recognize the object or environment it is interacting with, like the terrain type in our case, the action (gait) that gives rise to the experienced sensory stimulation needs to be considered. In addition we have shown that deployment of longer action contexts further improves the discrimination capabilities. Our approach demonstrates that the robot successfully engages the acquired sensorimotor knowledge to optimize its behavior by selecting appropriate gaits on different ground substrates.

Apart from serving as a model of SMCT, our work has also substantial application potential. Autonomous, perception-based, off-road navigation is a hot research topic in mobile robotics (e.g., [9]). Unlike traditional approaches that rely on passive long-distance perception using high resolution sensors, we have hinted at the potential of a radically different approach: terrain perception through active generation of sensory stimulation in a multimodal collection of low-resolution sensors (for learning eSMCs, 1 bit per sensory channel was used). Taking action-observation sequences into account and exploiting the robot's rich

body dynamics to simplify the structure of the sensory information, an advantageous transformation of the input space for classification can be achieved.

In the current study we have employed only proprioceptive and contact sensors. These have proven very effective in ground discrimination and, in conjunction with a simple one-step prediction of the best next action based on the current sensorimotor context, the robot could optimize its behavior. However, these sensors provide little information about the terrain beyond the robot's current location. Distal sensors (like infrared or vision), on the other hand, could provide information about future events that could likewise be exploited for perceptual categorization and further improvement of the behavior. A promising approach in this respect uses internal simulation in sensorimotor space to find action sequences that optimize the success of the agent with a longer temporal horizon [13, 19]. Alternatively, reinforcement learning algorithms could be employed. Traversability in general may be a suitable touchstone to compare different approaches to use sensorimotor information for controlling robots. This will be the direction of our future work.

Acknowledgement

This work was supported by the Sino-German research training group CINACS, DFG GRK 1247, `www.cinacs.org`, and by the EU project 'Extending sensorimotor contingencies to cognition - eSMCs', FP7-ICT-270212, `esmcs.eu`.

References

1. J.A. Bagnell, D. Bradley, D. Silver, B. Sofman, and A. Stentz. Learning for autonomous navigation. *Robotics & Automation Magazine, IEEE*, 17 (2):74–84, 2010.
2. R.D. Beer. The dynamics of active categorical perception in an evolved model agent. *Adaptive Behavior*, 11:209–243, 2003.
3. J. Dewey. The reflex arc concept in psychology. *Psychological Review*, 3:357–370, 1896.
4. A.K. Engel. Directive minds: how dynamics shapes cognition. In Stewart J., Gapenne O., and Di Paolo E. A., editors, *Enaction: Towards a New Paradigm for Cognitive Science*, pages 219–243. MIT Press, Cambridge, MA, 2011.
5. E.J. Gibson, G. Riccio, M.A. Schmuckler, T.A. Stoffregen, D. Rosenberg, and J. Taromina. Detection of traversability of surfaces by crawling and walking infants. *Journal of Experimental Psychology*, 13(4):533–544, 1987.
6. P. Giguere and G. Dudek. Clustering sensor data for autonomous terrain identification using time-dependency. *Autonomous Robots*, 26:171–186, 2009.
7. S. Harnad. *Handbook of categorization in cognitive science*, chapter Cognition is categorization. Elsevier, 2005.
8. M. Hoffmann and R. Pfeifer. *The Implications of Embodiment: Cognition and Communication*, chapter The implications of embodiment for behavior and cognition: animal and robotic case studies, pages 31–58. Imprint Academic, 2011.
9. L.D. Jackel, E. Krotkov, M. Perschbacher, J. Pippine, and C. Sullivan. The DARPA LAGR program: goals, challenges, methodology, and phase I results. *Journal of Field Robotics*, 23(11/12):945–973, 2006.

10. M. Lungarella and O. Sporns. Mapping information flow in sensorimotor networks. *PLoS Computational Biology*, 2 e144(10):1301–1312, 2006.
11. A. Maye and A.K. Engel. A discrete computational model of sensorimotor contingencies for object perception and control of behavior. In *2011 IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 3810–3815. IEEE, May 2011.
12. A. Maye and A.K. Engel. Time scales of sensorimotor contingencies. In H. et al. Zhang, editor, *Brain-Inspired Cognitive Systems 2012*, volume 7366 of *LNAI*, pages 240–249. Springer-Verlag Berlin, Heidelberg, 2012.
13. R. Möller and W. Schenck. Bootstrapping cognition from behavior – a computerized thought experiment. *Cognitive Science*, 32(3):504–542, 2008.
14. A. Noë. *Action in perception*. MIT Press, 2004.
15. J.K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24:939–1031, 2001.
16. R. Pfeifer and C. Scheier. Sensory-motor coordination: The metaphor and beyond. *Robotics and Autonomous Systems*, 20:157–178, 1997.
17. M. Reinstein and M. Hoffmann. Dead reckoning in a dynamic quadruped robot: Inertial navigation system aided by a legged odometer. In *Robotics and Automation (ICRA), 2011 IEEE Int. Conf. on*, pages 617–624, 2011.
18. A. Sirigu, J. R. Duhamel, and M. Poncet. The role of sensorimotor experience in object recognition. A case of multimodal agnosia. *Brain : a journal of neurology*, 114 (Pt 6):2555–2573, December 1991.
19. E. Ugur, E. Oztop, and E. Sahin. Goal emulation and planning in perceptual space using learned affordances. *Robotics and Autonomous Systems*, 59:580–595, 2011.
20. E. Ugur and E. Sahin. Traversability: a case study for learning and perceiving affordances in robots. *Adaptive Behavior*, 18:258–284, 2010.
21. Webots. www.cyberbotics.com. Commercial Mobile Robot Simulation Software.
22. D.A. Wolk, H.B. Coslett, and G. Glosser. The role of sensory-motor information in object recognition: Evidence from category-specific visual agnosia. *Brain and Language*, 94(2):131 – 146, 2005.

Appendix E

Actor-Critic Design using Echo State Networks in a Simulated Quadruped Robot

Schmidt, N. M., Baumgartner, M. and Pfeifer, R. (2014). "Actor-critic design using echo state networks in a simulated quadruped robot." In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014)*, pp.2224–2229, 14–18 Sept. 2014, Chicago, IL, USA, doi:10.1109/IROS.2014.6942862.

Actor-Critic Design using Echo State Networks in a Simulated Quadruped Robot

Nico M. Schmidt[1,*], Matthias Baumgartner[1] and Rolf Pfeifer[1]

Abstract— **In recent years, several studies proposed the application of echo state networks (ESN) to adaptive reinforcement learning schemes for the control of artificial autonomous agents. Especially the actor-critic design (ACD) is a promising candidate for robotic systems with continuous state and action spaces, as was demonstrated in several studies using simple wheeled robots. In the present work, we investigate applicability of this learning framework to more complex robotic systems, namely a quadruped running robot with rich dynamics. New challenges and questions arise, such as the nontrivial mapping of actions to the resulting behavior.**

I. INTRODUCTION

One of the biggest challenges in artificial intelligence is how artificial autonomous agents can learn a task or behavior – similar to their biological counterparts – from scratch with minimum prior knowledge about their body and environment. The classical approach is to create a model of the robot and its environment that can be used for control. This is infeasible for more complex robotic systems acting in the real world. Thus, it is desirable to rather let the agent learn a task from its own experience through active exploration of its capabilities. Reinforcement learning provides a promising framework for this purpose. The agent receives a feedback (reward) for its actions, which is often much easier to implement than – like in supervised learning – to teach the action that would have been the best choice (as it is often unknown to the supervisor either). Estimating the rewards of future time steps and using this estimate to select the most promising actions is the core of this approach. It has been pursuit in many ways and many scenarios during the last decades (see e. g. [1], [2]).

As the agents get more complex, most of the proposed methods suffer from the "curse of dimensionality": With the state and action spaces becoming larger, the computational complexity increases exponentially. This especially affects the otherwise promising "dynamic programming". This problem has been tackled with heuristic approaches using artificial neural networks, known as "adaptive critic designs" (ACD)[3] (also "actor-critic designs"). In ACD, the estimation of the action-value (or utility) function is the task of the critic element, which is separated from the actor network that selects the actions. The approximation of the utility function in form of the discounted sum of future rewards (expressed by the Bellman equation) is bootstrapped iteratively, using in each step the current estimate together with the reward for the previously taken action.

ACD methods require approximation techniques which are (a) able of modeling the highly nonlinear and dynamic systems constituted by robots operating in real-time in uncertain environments and (b) efficient in online training. A relatively novel machine learning technique that fulfills these requirements are echo state networks (ESN) [4]. They recently gained wide attention due to their outstanding performance in many real-world applications. ESNs are recurrent neural networks that consist of a "reservoir"-part of recurrently connected neurons and a readout-part that connects the reservoir neurons to an output. The efficiency in training comes from the fact that only the readout weights are trained (e. g. by linear regression). The reservoir weights remain fixed while still providing complexity due to recurrent connections.

Some recent studies have demonstrated how ESNs can be successfully applied in an ACD to estimate the action-value function in order to control a two-wheeled robot avoiding obstacles [5], [6], or moving towards reward regions and away from punishment regions [7].

While the mapping of action space to resulting behavior is straightforward in the case of a two-wheeled robot, the question remains how the ESN-ACD method copes with systems having more rich dynamics, where action-behavior mappings are much more complex and depend on a larger history. In the present paper, we attempt to answer this question by applying the method to an under-actuated compliant quadruped robot performing a navigational task. We extend the proposed actor network to meet the new requirements arising in this setup.

II. METHODS

A. The Learning Problem

The agent-environment interaction is formalized by the transition of the agent's state from $\vec{s}_t$ to $\vec{s}_{t+1}$ while executing action $\vec{a}_t$. At each time step, the agent receives a reward r_t for its previously chosen action. The agent's goal is then to choose the actions that maximize the utility function, which is calculated as the discounted sum of future rewards:

$$J_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \tag{1}$$

The discount rate $\gamma \in [0, 1]$ weights rewards of the near future higher than rewards that are still far away. The task of the critic is to estimate J, while the actor has to use this estimate to select $\vec{a}_{t+1}$ such that J is maximized.

The approximation $\hat{J}(\vec{s}_t, \vec{a}_t)$ is trained in a supervised way, according to a bootstrapping version of the *SARSA*

[1]The authors are with the Artificial Intelligence Laboratory, Dept. of Informatics, University of Zurich, 8050 Zurich, Switzerland

*Corresponding author, nico.schmidt@uzh.ch. This work was supported by the EU FP7 project Extending Sensorimotor Contingencies to Cognition (eSMCs), IST-270212.

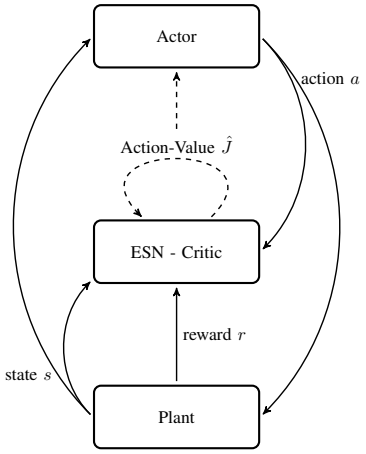


Fig. 1. Actor-Critic schema. The actor outputs an action and is criticized for its choice by the critic. Dashed lines indicate learning signals. The critic uses reward, action and its estimate to calculate the TD-error and adjust its parameters. The actor uses the current estimate $\hat{J}$ to chose the next action.

algorithm [2]. As in *SARSA*, the temporal-difference (TD) error is minimized iteratively, using in each step the estimate and reward of the previous step:

$$E_{TD} = r_{t+1} + \gamma \hat{J}(\vec{s}_{t+1}, \vec{a}_{t+1}) - \hat{J}(\vec{s}_t, \vec{a}_t) \tag{2}$$

The actor chooses the action $\vec{a}_{t+1}$ based on the critic's current estimate. While in the beginning of the training the estimates might lead to sub-optimal actions, they will approach the optimal $\vec{a}^*_{t+1}$ as $\hat{J}$ approaches J. Fig. 1 shows a schematic of the ACD architecture.

Through actions and resulting state transitions, the utility function is dependent on the nonlinear body dynamics of the robot. Because ESNs are universal approximators of dynamical systems and thus promising candidates for capturing those dynamics, we use an ESN to estimate J (Fig. 2). The reservoir part with activation $\vec{x}_t$ is driven by the input through $\mathbf{W}^{\text{in}}$ and its previous activation through $\mathbf{W}^{\text{res}}$:

$$\vec{x}_t = f(\mathbf{W}^{\text{in}} \vec{u}_t + \mathbf{W}^{\text{res}} \vec{x}_{t-1}) \tag{3}$$

where $\vec{u}_t$ is the input vector, a concatenation of the state and action vectors, $\vec{u}_t = \begin{bmatrix} \vec{s}_t & \vec{a}_t \end{bmatrix}$. For the activation function f, the hyperbolic tangent $\tanh$ is used. The predicted utility $\hat{J}$ is calculated as a linear combination of reservoir activations and inputs:

$$\hat{J}_t = \mathbf{W}^{\text{out}} \begin{bmatrix} \vec{u}_t \\ \vec{x}_t \end{bmatrix} \tag{4}$$

The weights $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{res}}$ are randomly chosen and remain fixed, while the output weight $\mathbf{W}^{\text{out}}$ are trained iteratively using a recursive least squares (RLS) algorithm [8, p. 423], using the TD-error (2) as teaching signal. The most important free parameters are the scaling of the input, the spectral radius of the reservoir weights, the number of neurons in the reservoir and the connectivity density of $\mathbf{W}^{\text{in}}$ and $\mathbf{W}^{\text{res}}$.

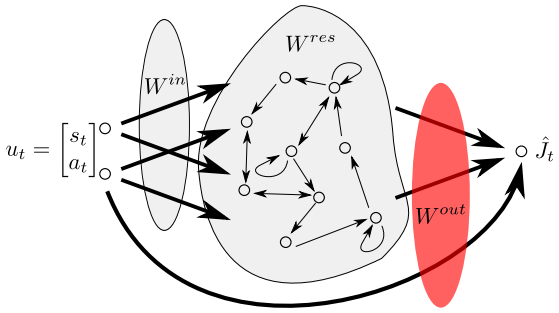


Fig. 2. Echo state network. The input weights and the recurrent reservoir weights are randomly initialized and remain fixed. Only the output weights are trained by a regression method.

The actor uses the ESN-critic to select the next action. The idea is to use gradient ascent on the utility function with respect to the previous action $\frac{\partial \hat{J}}{\partial \vec{a}}$. The gradient of an ESN's output w. r. t. its input depends only on the input and output weights, as well as on the current activations of the reservoir neurons. It is not necessary to account for the recursive structure of the reservoir, which makes the computation very efficient compared to e. g. backpropagation through time. For the gradient ascent, we repeat several (G) iterations, each time calculating the gradient of the utility function at the current action and updating the action according to

$$\begin{aligned} &\vec{a}' \leftarrow \vec{a}_t \\ &\textbf{for } k = 1 \dots G \textbf{ do} \\ &\quad \vec{a}' \leftarrow \vec{a}' + \eta \frac{\partial \hat{J}_t}{\partial \vec{a}'} \\ &\textbf{end} \\ &\vec{a}_{t+1} \leftarrow \vec{a}' \end{aligned} \tag{5}$$

It is assumed that the input matrix is composed of two sub-matrices, one for the state input and one for the action input $\mathbf{W}^{\text{in}} = \begin{bmatrix} \mathbf{W}^{\text{in}}_s & \mathbf{W}^{\text{in}}_a \end{bmatrix}$. Similarly, the output weight matrix is assembled from the reservoir states, state input and action input matrices $\mathbf{W}^{\text{out}} = \begin{bmatrix} \mathbf{W}^{\text{out}}_{x,J} & \mathbf{W}^{\text{out}}_{s,J} & \mathbf{W}^{\text{out}}_{a,J} \end{bmatrix}$. The gradient is then obtained using the chain rule as follows.

$$\frac{\partial \hat{J}_t}{\partial \vec{a}_t} = \underbrace{\frac{\partial \hat{J}_t}{\partial \vec{x}_t}}_{\mathbf{W}^{\text{out}}_{x,J}} \cdot \underbrace{\frac{\partial \vec{x}_t}{\partial f}}_{(\mathbb{1} - \vec{x}_t^2)} \cdot \underbrace{\frac{\partial f}{\partial \vec{a}_t}}_{\mathbf{W}^{\text{in}}_a} + \underbrace{\frac{\partial \hat{J}_t}{\partial \vec{a}_t}}_{\mathbf{W}^{\text{out}}_{a,J}} \tag{6}$$

B. Robot Simulation

The simulated robot used in the present work is a computer model of the quadruped Puppy robot (Fig. 3), whose mechanical design (weight distribution, proportions, springs used, etc.) is a result of previous research (e.g. [9]). The robot has four identical legs driven by position-controlled servomotors in the hips. It has passive compliant joints at the knees. Upper and lower limbs are connected with springs. The computer model[1] was designed in Webots [10], a physics-

[1]The source code of the Puppy model (`https://github.com/eSMCs/PuPy`) and the Python ESN-ACD implementation (`https://github.com/eSMCs/HDPy`) are freely available.

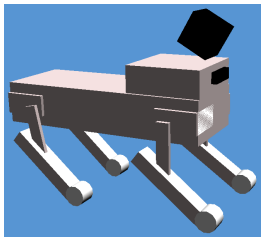


Fig. 3. The Puppy robot model in the Webots simulator.

TABLE I
THE PARAMETERS USED IN THE EXPERIMENTS.

Echo state network settings	
#reservoir neurons N	100
spectral radius σ	0.7
connection density of $\mathbf{W}^{\text{res}}$	100%
connection density of $\mathbf{W}^{\text{in}}$	20%
Actor-Critic settings	
discount factor γ	0.5
action step rate η	0.1
random action rate ϵ_{greedy}	0.05
#gradient iterations G	50

based simulation software.

The four motors $i \in \{FL, FR, HL, HR\}$ (F=front, H=hind, L=left, R=right) are each controlled using a sine wave pattern generator:

$$\text{pos}_i(t) = \alpha_i \cdot \sin(2\pi\omega t + \theta_i) + \beta_i \tag{7}$$

with adjustable parameters amplitude α, frequency ω, offset β and phase shift θ.

C. Experimental Settings

The robot's task was a navigational one, where the robot is rewarded for walking towards a target position at $(-1000, 1000)$ (direction of south-east). To be able to sense its orientation, the robot was equipped with a compass sensor, giving a 3-dimensional unit vector pointing to the virtual north, $\vec{s}_t = [c_x, c_y, c_z]^T + \epsilon_{\text{noise}}$ with input noise $\epsilon_{\text{noise}} \sim \mathcal{N}(0, 0.05)$. The reward function we used punished the deviation of the robot's heading from the direction of the target. Furthermore, the robot received a punishment for falling over:

$$r_t = 1 - \frac{\phi}{\pi} - \delta_{\text{tumble}} \tag{8}$$

with ϕ being the absolute angle between the robots heading direction and the direction towards the target. δ_{tumble} is 1 if the robot tumbled during the last control period, otherwise 0. Thus, the rewards lay in the range $[-2, 0]$.

For control, we used the sine wave parameters of a bounding gait[2], learned in a previous study [11]. With these bounding gait parameters the robot runs straight ahead. To enable the robot to navigate left- and rightwards, we let the actor vary the sine amplitudes of the two left ($\alpha_L := \alpha_{FL} = \alpha_{HL}$) and the two right legs ($\alpha_R := \alpha_{FR} = \alpha_{HR}$) at every two-seconds control step. This resulted in a two-dimensional action space: $\vec{a}_t = [\alpha_L, \alpha_R]$.

The simulation was run in episodes, which ended when the robot tumbled or ran out of the arena (50x50 meters). In the beginning of each episode the robot was placed in the arena at a random position with random heading. For the first 20 control steps the robot ran with randomly chosen amplitudes in the range $[0.2, 0.9]$ to wash out the reservoir activations. After the 20^{th} step the actor took over control and adjusted the action every two seconds using the gradient method in (5). However, an ϵ-greedy policy was implemented, such that a random action was selected with probability ϵ_{greedy} and the actor's choice with probability $(1 - \epsilon_{\text{greedy}})$. In order to guarantee enough variety in the selected actions, that is necessary for robust training of the critic, the parameter ϵ_{greedy} was set to 0.8 in the beginning and slowly decreased to 0.05 over time. Furthermore, the discount rate γ was set to 0 in the beginning and slowly increased in each episode up to a value of 0.5, such that the robot's time horizon expanded over time.

Table I lists the parameters used in the experiments.

III. RESULTS

A. Action-Behavior Mapping

Intuitively, we can roughly assume that larger amplitudes on the left legs than on the right legs ($\alpha_L > \alpha_R$) lead to a rightward turning and vice versa. The actual mapping, however, is much more complex and not unique. The change in heading largely depends on previous actions and current posture of the robot. The prediction of the behavior (and the utility J) is thus a difficult task and requires a certain memory of previous states and actions. An example of behaviors resulting from different actions is shown in Fig. 4. Depicted are the behaviors (expressed as change in heading) for actions selected at $t = 21$ after running straight ahead for 20 control steps ($\vec{a}_t = [0.5, 0.5]$ for $t = 1, ..., 20$). Note that this action-behavior map is specific for the action history (and thus state-history) and would look different if different actions would have been selected before. The figure shows that actions leading to sharp turning bear higher risk to tumble. Thus the robot has to learn a trade-off between turning speed (i. e. reaching the target heading in fewer steps) and stable walking.

B. Online Learning

Fig. 5 summarizes the convergence of the critic's online training. It can be seen that the magnitude of the weight updates (bottom panel) is highest in the first episodes and approaches a constant magnitude of about 10^{-4} after approx. 8 hours. In the beginning, the robot tumbled quite often (as can be seen by the short episode lengths), while it manages to run stable for most of the time after the weights converged.

The TD-error (top panel) shows occasional peaks, mostly at short episodes, i. e. when the robot tumbled. This indicates that during tumbling the utility J is more difficult to predict

[2]Gait parameters: $\omega = 1$Hz, $\beta_{FL/FR} = -0.23$, $\beta_{HL/HR} = -0.2$, $\theta_{FR/FL} = 0$, $\theta_{HR/HL} = 0.5$, $\alpha_{FR/FL} = 0.56$, $\alpha_{HR/HL} = 0.65$

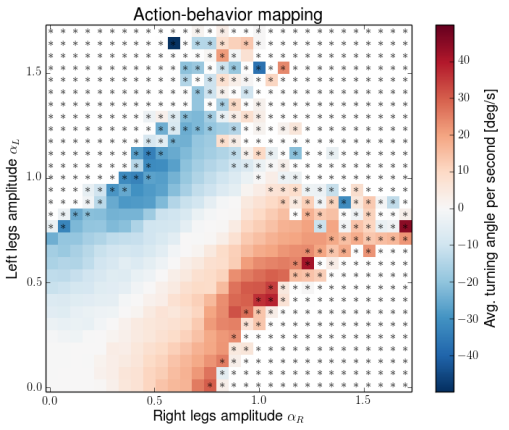


Fig. 4. Action-behavior mapping for actions performed after running straight ahead for 20 control steps ($\vec{a}_{t=1\ldots20} = [0.5, 0.5]$). The color indicates the turning angle of the robot when performing the respective action in the next control step $\vec{a}_{t=21} = [\alpha_L, \alpha_R]$. Rightwards turning is indicated as blue colors, leftwards turning as red colors. Actions that lead to tumbling of the robot are marked with an asterisk character.

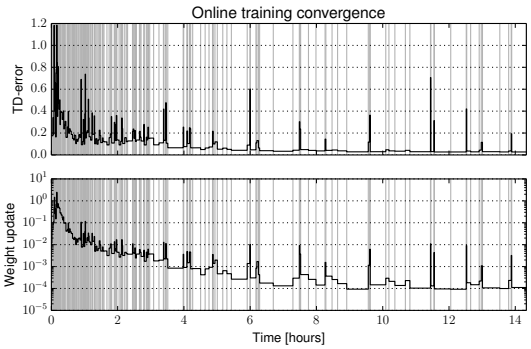


Fig. 5. TD-error (top) and output weights (bottom) over time. The absolute values of the TD-error according to (2) is shown (average within episodes). The bottom plot shows the mean of the absolute weight update $|\Delta \mathbf{W}^{\text{out}}|/(n_{\text{rows}} + n_{\text{columns}})$, also averaged within episodes. The vertical gray lines indicate the episode starts. Note the log scale in the bottom panel.

than during the rest of the episodes. The reason might be that tumbles occur relatively rarely, hence these input samples are underrepresented in the data set. Moreover, the input states consist of compass sensors only, which might not carry enough information for robust tumbling detection. Adding acceleration sensors or joint angle sensors might improve the critic in this respect.

Fig. 6 compares the performance of the critic and the actor before convergence of the training ($t < 2.5$ hours, top panels) and after convergence ($t > 8$ hours, bottom panels). The upper plot in the left column confirms that the critic was able to map the compass sensors to the utility function correctly already in the first 2.5 hours of learning. The predicted utility was highest when Puppy headed south-east, where the target was located (red colors), whereas it was low when Puppy ran into another direction (green and blue colors). However, the actor was not yet able to select the right actions. The preferred heading direction, indicated in the upper plot of the middle column, is not showing any clear preference and shows all possible headings occurring with somewhat more frequent occurrence of south and east directions. Also the trajectories in the upper right plot show rather short runs in random directions, always ending with tumbling of the robot and receiving of negative rewards (indicated by the blue colors in the end of the trajectories).

After 8 hours of training and convergence of the output weights, the critic was not only able to approximate the utility function, but furthermore the actor was able to correctly make use of this knowledge in selecting the actions that navigated Puppy towards the target location. The prediction is better distinguishing the compass values (lower left plot) and the robot shows a clear preference to walk towards south-east (lower plot of the middle column). The trajectories in the lower right plot all show movements towards the target direction (south-east) and are much longer, meaning the robot did not tumble any more.

The fact that at 2.5 hours the critic showed reasonable utility estimates, but the actor was not able to select the right actions, indicates that the mapping of the compass values to the reward was captured correctly, but the effect of tumbling was not yet well enough represented by the estimate. Recalling Fig. 4, this might be due to the sharp edge between actions that lead to turning and actions that lead to tumbling. At this time, the estimated utility function might have been too smooth, predicting the right combination of amplitudes for turning, but not taking into account the tumbling that occurred with too high amplitudes.

C. Gradient Ascent

Having a more detailed look at the learned utility function reveals how the gradient ascent is performing. Fig. 7 shows the action selection process using gradient ascent after running for 30 control steps with action $\vec{a}_{t=1\ldots30} = [0.5, 0.625]$ (leading to a rightwards turning). Depicted is the resulting reward for all possible follow-up actions $\vec{a}_{t=31}$ (upper plot), and the respective estimated utility $\hat{J}_{t=31}$ (lower plot). The actions resulting in a maximum reward (red colors) are the ones close to the "tumbling-edge" (dark blue). The estimated utility captures the overall turning effect of the actions, but almost ignores the tumbling regions. Note that the current estimate is specific to the history of actions and states (represented in the reservoir activations). It appears that the estimate and thus the gradient is accurate only locally around the previously performed action. Too many gradient iterations might therefore lead to regions of the action space where the estimate is poor. By adjusting the number of gradient iterations, this behavior can be balanced: on the one hand, only few iterations are not sufficient to reach high-rewarded actions, on the other hand, too many iterations lead to uncertain outcomes.

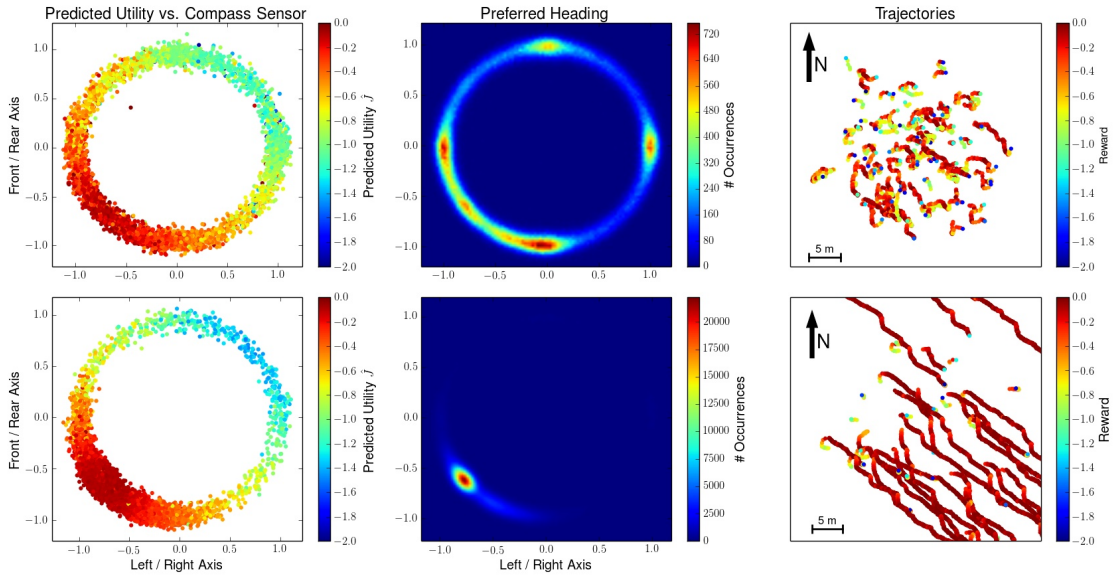


Fig. 6. ACD performance before and after convergence. The three plots at the top show the training results for the first 100 episodes (2.5 hours) and the bottom plots show the results after the 150^{th} episode (8 hours). The predicted utility is plotted on the left as a function of the compass sensor values (input states $\tilde{s}$). The location of the samples indicates the current compass values, the utility values are indicated by the color. The plots in the middle column show how often each compass value has been measured. The trajectories of the robot in the simulated world are shown in the right plots, with the actual rewards for each time step indicated by the colors.

IV. DISCUSSION

In the present study, we used an ESN as approximator of the utility function in an actor-critic reinforcement learning scheme on a simulated quadruped robot. Other than in previous studies, where the system was demonstrated on two-wheeled robots, the situation is much more difficult in a quadruped one. Due to the complex dynamics of the Puppy robot, the state transitions mediated by the chosen actions are highly nonlinear and dependent on the robot's state-action history. A prediction of future rewards therefore needs an approximator which is able to capture those dynamics. We demonstrated that the ESN-critic can cope with these difficulties. Besides online training capability, its architecture allows for fast derivative calculation, enabling the gradient ascent-based action selection above real-time[3].

One issue that seems to affect learning performance a lot is that of the action sampling strategy. For the critic to be reliable, it must have seen many observations from many different possible state-action histories. Thus, actions need to be sampled from a broad but viable range of the action space. In our experiments, we use the parameter ϵ_{greedy} to control the probability of random action selection. Our experiments revealed that using low ϵ_{greedy} from the beginning leads to a fast convergence of the training, but results in a poor representation of the utility function (data not shown here). Following the suggestion from previous studies [6], [5] to start from a high ϵ_{greedy} and slowly decreasing it to small values, proved to result in more robust control in the experiments presented here, while drastically increasing the time needed for training. When applying the method to real physical robots, this trade-off needs to be taken into account. Our results suggest that with a suitable choice of ϵ_{greedy}, the method could well be applied to the real Puppy robot. Instead of selecting random actions in the nongreedy control steps, more sophisticated action sampling strategies could be employed, e. g. goal directed ones such as online goal-babbling [12], [13], or uncertainty-based methods that select the actions to minimize the uncertainty of the estimate using e. g. Gaussian Process Regression [14].

In our setting, the actor was selecting actions by gradient ascending the utility function starting from the previously selected action. The possibility of more sophisticated actor networks remains to be investigated. Interesting candidates are the selection of actions through additional readout connections from the ESN, which could be used to produce the next action based on the current reservoir activations. The weights of such connections could be trained with a similar gradient method as was used here. Other approaches, such as "reward-modulated Hebbian learning" (see e. g. [15], [16]) could be employed, where the reward is used as a step rate for adjusting the output weights during Hebbian learning.

Further investigations are needed towards a deeper understanding of the relation between the history of action-outcome dynamics and the reservoir memory capacity. To

[3]The simulations ran in 3-5x real-time using a Python implementation of the ACD on a 2.6GHz quad-core personal computer.

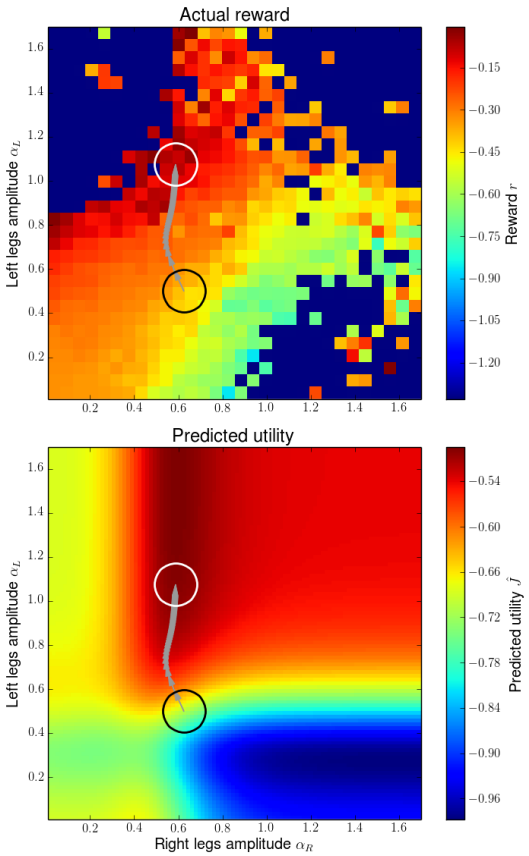


Fig. 7. Gradient ascending the utility function. Actual reward (top) and estimated utility (bottom) for all combinations of left and right amplitudes (α_L and α_R). Depicted is the situation after running 30 steps with action $\vec{a}_{t=1\ldots30} = [0.5, 0.625]$. At this time, the target lies in the front-right direction of the robot, thus a rightward movement is desired. The gray arrows indicate the gradient steps, leading from $\vec{a}_{30}$ (black circle) to $\vec{a}_{31}$ (white circle).

this end, the influence of the reservoir parameters on the critic's and actor's performance needs systematic investigation. Especially the spectral radius and number of reservoir neurons are likely to influence the performance, since they largely determine the memory capacity of the ESN.

Other studies suggested further, the usage of intrinsic plasticity [17] to adapt the reservoir activation function to the input characteristics. Especially in practical applications this has been proven to increase prediction performance significantly [18].

In conclusion, the ESN-ACD method is a promising technique for artificial agents that are to learn tasks from scratch in a bootstrapping way, even for complex robots with rich dynamics such as the Puppy robot. Future directions of this work will include systematic parameter studies, investigation of increasing task-complexity and more sophisticated actor-networks. Finally, the applicability to a real physical robot remains to be shown.

References

[1] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement Learning in Robotics : A Survey," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[2] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA, 1998 A Bradford Book, 1998.

[3] D. V. Prokhorov and D. C. Wunsch, "Adaptive critic designs." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 8, no. 5, pp. 997–1007, Jan. 1997.

[4] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note'," Tech. Rep., 2001.

[5] P. Koprinkova-Hristova, M. Oubbati, and G. Palm, "Heuristic dynamic programming using echo state network as online trainable adaptive critic," *International Journal of Adaptive Control and Signal Processing*, vol. 27, no. 10, pp. 902–914, 2012.

[6] M. Oubbati, M. Kächele, P. Koprinkova-Hristova, and G. Palm, "Anticipating rewards in continuous time and space with echo state networks and actor-critic design," *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, no. April, pp. 27–29, 2011.

[7] M. Oubbati, J. Uhlemann, and G. Palm, "Adaptive Learning in Continuous Environment Using Actor-Critic Design and Echo-State Networks," *From Animals to Animats 12*, pp. 320–329, 2012.

[8] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., Hoboken, New Jersey, 1998.

[9] F. Iida, G. Gómez, and R. Pfeifer, "Exploiting body dynamics for controlling a running quadruped robot," in *Proceedings of the 12th International Conference on Advanced Robotics ICAR2005*. Seattle, WA, USA: IEEE, 2005, pp. 229–235.

[10] O. Michel, "Webots: Professional Mobile Robot Simulation," *Journal of Advanced Robotics Systems*, vol. 1, no. 1, pp. 39–42, 2004.

[11] M. Hoffmann, N. M. Schmidt, K. Nakajima, F. Iida, and R. Pfeifer, "Perception, motor learning, and speed adaptation exploiting body dynamics: case studies in a quadruped robot," in *Proceedings of the 5th International Symposium on Adaptive Motion of Animals and Machines (AMAM 2011)*, 2011, pp. 39–40.

[12] M. Rolf, J. J. Steil, and M. Gienger, "Online Goal Babbling for rapid bootstrapping of inverse models in high dimensions," in *2011 IEEE International Conference on Development and Learning (ICDL)*. Ieee, Aug. 2011, pp. 1–8.

[13] C. Hartmann, J. Boedecker, O. Obst, S. Ikemoto, and M. Asada, "Real-Time Inverse Dynamics Learning for Musculoskeletal Robots based on Echo State Gaussian Process Regression," in *Proceedings of Robotics: Science and Systems VIII*, N. Roy, P. Newman, and S. Srinivasa, Eds. Sydney, Australia: MIT Press, 2012.

[14] P. a. Romero, A. Krause, and F. H. Arnold, "Navigating the protein fitness landscape with Gaussian processes." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. E193–201, 2013.

[15] R. Legenstein, S. M. Chase, A. B. Schwartz, and W. Maass, "A reward-modulated hebbian learning rule can explain experimentally observed network reorganization in a brain control task." *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 30, no. 25, pp. 8400–10, June 2010.

[16] T. McMillen, P. Simen, and S. Behseta, "Hebbian learning in linear-nonlinear networks with tuning curves leads to near-optimal, multi-alternative decision making." *Neural networks : the official journal of the International Neural Network Society*, vol. 24, no. 5, pp. 417–26, June 2011.

[17] J. Triesch, "A Gradient Rule for the Plasticity of a Neurons Intrinsic Excitability," in *15th International Conference on Artificial Neural Networks (ICANN 2005)*, 2005, pp. 1–7.

[18] P. Koprinkova-Hristova, M. Oubbati, and G. Palm, "Adaptive Critic Design with Echo State Network," *2010 IEEE International Conference on Systems, Man and Cybernetics*, pp. 1010–1015, Oct. 2010.

Appendix F

Learning and Adaptation of Sensorimotor Contingencies: Prism-Adaptation, a Case Study

Reprinted from:

Kootstra, G., Wilming, N., Schmidt, N. M., Djurfeldt, M., Kragic, D. and König, P. (2012). "Learning and Adaptation of Sensorimotor Contingencies: Prism-Adaptation, a Case Study". In Ziemke, T., Balkenius, C. and Hallam, J. (Eds.), *From animals to animats 12: Proc. Int. Conf. Simulation of Adaptive Behavior (SAB)*, Odense, Denmark, Vol. 7246 of LNAI, Springer, pages 341–350, doi:10.1007/978-3-642-33093-3_34.

This is the final accepted version. The original publication is available at www.springerlink.com (http://link.springer.com/chapter/10.1007/978-3-642-33093-3_34).

Learning and Adaptation of Sensorimotor Contingencies: Prism-Adaptation, a Case Study

Gert Kootstra[1], Niklas Wilming[2], Nico Schmidt[3], Mikael Djurfeldt[4], Danica Kragic[1] and Peter König[5]

[1] CAS-CVAP, CSC, Royal Institute of Technology (KTH). {kootstra,dani}@kth.se
[2] Institute of Cognitive Science, University of Osnabrück. nwilming@uos.de
[3] AI Lab, University of Zürich. nico.schmidt@uzh.ch
[4] PDC, Royal Institute of Technology (KTH). mdj@kth.se
[5] Institute of Cognitive Science, University of Osnabrück; Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf. pkoenig@uos.de

Abstract. This paper focuses on learning and adaptation of sensorimotor contingencies. As a specific case, we investigate the application of prism glasses, which change visual-motor contingencies. After an initial disruption of sensorimotor coordination, humans quickly adapt. However, scope and generalization of that adaptation is highly dependent on the type of feedback and exhibits markedly different degrees of generalization. We apply a model with a specific interaction of forward and inverse models to a robotic setup and subject it to the identical experiments that have been used on previous human psychophysical studies. Our model demonstrates both locally specific adaptation and global generalization in accordance with the psychophysical experiments. These results emphasize the role of the motor system for sensory processes and open an avenue to improve on sensorimotor processing.

Keywords: Sensorimotor contingencies, prism-adaptation, motor learning/adaptation, body maps, inverse kinematics

1 Introduction

Humans adapt easily to changes in sensorimotor coordination during development and adulthood. A remarkable demonstration is adaptation to prism glasses that displace the visual field horizontally by a constant angle. Despite such drastic changes of a sensorimotor relationship, eye-hand coordination quickly adapts [6].

This is, however, not a passive process but requires active exploration [4, 5]. Prism adaptation is specific to the involved body parts [7, 15] and actions [1, 7, 16]. The adaptation to changed sensorimotor dependencies therefore appears to crucially depend on the ability to relate one's own motor actions to their observed sensory consequences. This bears a striking resemblance to the concept of sensorimotor contingencies (SMC) [11], which also stresses the importance of action for perception [3]. The case of prism adaptation is well investigated and therefore may serve as a prime example for investigating how the brain achieves "mastery of a sensorimotor contingency" [11].

When prisms are donned, subject's pointing movements are offset due to the visual displacement. However, with repeated movements this offset diminishes and original

performance restores. When subsequently the prisms are removed, pointing movements are offset in the opposite direction. This *aftereffect*, i.e. the difference in pointing between pre- and post-exposure, is a convenient measure of the degree of adaptation.

The adaptation is thought to combine two separate processes: Recalibration and realignment [15]. Recalibration is believed to utilize a cognitive learning strategy that quickly reduces pointing errors. The effects of recalibration are local to specific actions. Recalibration is quantified by the aftereffect measured with the identical movement that was carried out during the prism exposure phase [15]. Realignment, in contrast, is thought to be an automatic process that aligns, for example, visual and proprioceptive maps. It reveals a global generalization of adaptation to new pointing targets and actions. Realignment is measured by actions not practiced during prism exposure.

Redding and Wallace observe that the realignment effect is modulated by different kinds of feedback during exposure. When participants could see their own hand-movement (concurrent feedback) during exposure, Redding and Wallace [15] observed a small shift of "visual straight ahead" and a large shift of "proprioceptive straight ahead". This pattern was reversed when participants could only see the end-position of their hand (terminal feedback) during exposure. Related to this, Redding and Wallace found that the aftereffect generalized differently to targets not shown during exposure. Specifically, concurrent feedback produced aftereffects that increased for targets in the direction of the prismatic shift, whereas they decreased for terminal feedback. Both results are explained by two different references: During concurrent feedback the visual system acts as a reference and the proprioceptive system is aligned to it and during terminal feedback the situation is reversed.

In this paper, we investigate properties of recalibration and realignment and their dependence on different types of feedback. For this purpose, a computational model of prism adaptation is developed for the control of a simulated robotic arm and subjected to the identical experiments focused on eye-hand coordination of the previous psychophysical study [15]. Specifically, we test the hypothesis that the different forms of adaptation aftereffects can be described in a unified approach based on the concept of sensorimotor contingencies. We take as our starting point that adaptive agents have to relate changes in perception to their own actions. At the same time we strive for a computational description of recalibration and realignment to foster our understanding of sensorimotor adaption and for facilitating the development of versatile robotic agents. The ability to quickly learn eye-hand coordination from own experience makes manual calibrations redundant, which is interesting, for instance, for robotic grasping [12].

2 Computational model of prism adaptation

The contingency between sensory and motor signals involved in pointing movements is captured by forward and inverse-kinematic models. The forward model defines the position of the effector, e.g. the hand, based on the joint angles of the arm. The inverse-kinematic model provides the joint configuration of the arm necessary to reach a specific target, and thereby allows direct control of a robot. Motivated by [2, 10] and in accordance with the SMC theory [11] we use a differential inverse-kinematic model that generates the action necessary to reach a desired *change* in visual position of the

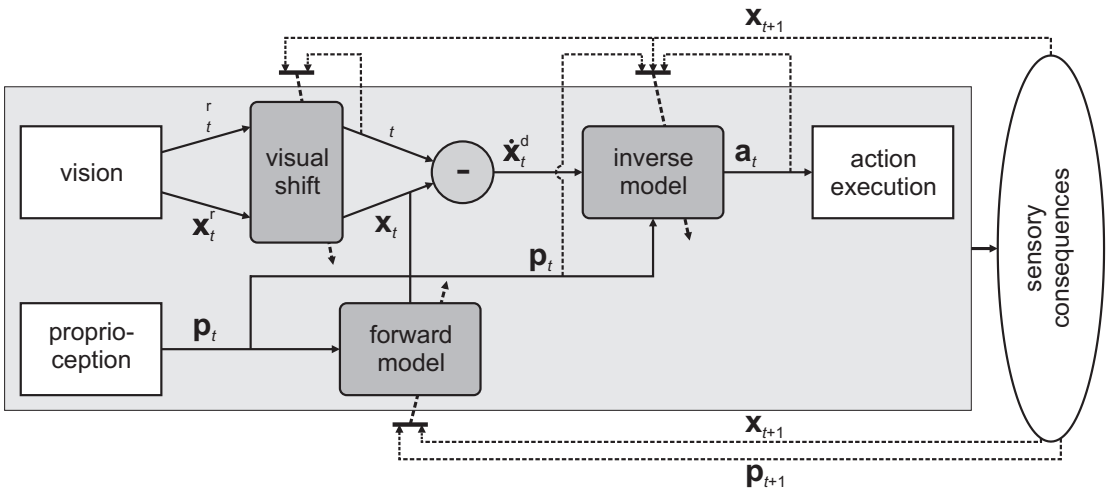


Fig. 1. Theoretical model, including three sub-modules: the visual shift, the forward model, and the inverse model. Based on visual and proprioceptive observations, the model estimates the action to get the end effector to the target pose. All three sub-modules learn on-line using the sensory consequences of the executed action as observed by the system itself. Information flow for control is indicated by solid arrows, and the flow for learning by dashed arrows. The variables are explained in the text.

end-effector. Given a current joint configuration, $\mathbf{q}$, and a desired *change* in pose of the end effector, $\dot{\mathbf{x}}$, the necessary *change* in joint angles, $\dot{\mathbf{q}}$, is determined by the model, expressed by the mapping $(\mathbf{q}, \dot{\mathbf{x}}) \mapsto \dot{\mathbf{q}}$. For small enough changes, the optimal gradient $\dot{\mathbf{q}}$ for any given state $\mathbf{q}$ is well defined and resolves ambiguities inherent in inverse-kinematics.

When the robotic system is subject to changes, for instance, due to mechanical wear or damage or adaptation to prisms, the forward and inverse mappings are not fixed, but changes over time. This calls for adaptive systems that can learn and adapt the inverse-kinematics model from own experience [9]. Others have used, for instance, locally-weighted projection regression [2, 17], Gaussian-process regression (GPR) [13], and local Gaussian-process regression [10]. Here we use Gaussian-process regression (GPR) to learn the inverse kinematics, since it has the best reported performance [13, 10] at the cost of higher computational load. But note, that the GPR can be easily substituted by other regression methods, if real-time performance is more of an issue.

In total, our model consists of three sub-modules: the forward model, the inverse model, and the visual-shift model (see Figure 1). In the following, we give the model details of all components in turn.

2.1 Forward model and inverse model

Based on the proprioceptive information $\mathbf{p}_t$, the *forward model* gives an estimation of the pose of the end effector in visual coordinates $\mathbf{x}_t$ at time t (see Figure 1). This estimation is used when the end effector cannot be observed visually. The *inverse model* provides an estimate of the necessary action, $\mathbf{a}_t$, for instance a change in joint angles,

based on $\mathbf{p}_t$ and the desired change in pose, $\dot{\mathbf{x}}_t^{\mathbf{d}}$, which is determined based on the difference between the target and the end effector, $\dot{\mathbf{x}}_t^{\mathbf{d}} = \tau_t - \mathbf{x}_t$ (see Figure 1).

Using GPR, an estimation is made of the functions $\mathbf{a}_t = f(\mathbf{p}_t, \dot{\mathbf{x}}_t^{\mathbf{d}})$ for the inverse and $\mathbf{x}_t = g(\mathbf{p}_t)$ for the forward model, based on a set of training examples, which the system acquires from own experience. GPR has a few hyperparameters, including $\{\lambda_1^2, \ldots, \lambda_D^2\}$, which are the characteristic length-scales of the different dimensions of the squared-exponential kernel, where D is the dimensionality of the input to the GPR. The hyperparameters are learned from data by maximizing the marginal likelihood. For details on GPR, we refer to [13].

Recency effect A problem with a standard GPR implementation of the forward and inverse models is that adaptation to changed SMCs is slow. When prism glasses are donned, the old training samples contribute as strong to the estimation as the new samples, which results in a rather slow adaptation. To increase speed of adaptation and to intensify the aftereffect, we include a forgetting mechanism using a recency effect, such that more recent training data make a stronger contribution to the estimation. Specifically, we add a time dimension to the input, resulting in input $\mathbf{z}^{\mathrm{i}} = \{\mathbf{p}, \dot{\mathbf{x}}, t\}$ for the inverse, and $\mathbf{z}^{\mathrm{f}} = \{\mathbf{p}, t\}$ for the forward model, both with a time constant λ_T. This characteristic length-scale parameter is not included in the optimizing of the hyperparameters, but instead used as a free parameter to control the recency effect, which is strong for low values of λ_T, whereas for $\lambda_T \to \infty$, the effect is absent.

Execution and on-line learning The Gaussian process regressors that implement the inverse and forward models are continuously updated while the robot performs its task, thus on-line learning the eye-hand coordination. In execution, the GPRs are used to predict the output of the models based on the SMC experience, resulting in an action. When the robot executes this action, it results in a movement of the arm. The system then observes the visual and proprioceptive consequences of that action and uses the observation of these new SMCs as training samples. In case of the forward model, the training data is $\{\mathbf{p}_{t+1}, \mathbf{x}_{t+1}, t\}$. The training data for the inverse model is $\{\mathbf{p}_t, \dot{\mathbf{x}}_t^{\mathbf{o}}, \mathbf{a}_t, t\}$, where $\dot{\mathbf{x}}_t^{\mathbf{o}} = \mathbf{x}_{t+1} - \mathbf{x}_t$ is the observed change in end effector.

2.2 Visual-shift model

The visual system provides information about the pose of the end effector and the target in visual coordinates. The visual-shift model applies a transformation, T, to these visual observations, so that the retinotopic (or camera) coordinate system is transformed into an internal visual coordinate system: $\mathbf{x} = T(\mathbf{x}^{\mathrm{r}})$ and $\tau = T(\tau^{\mathrm{r}})$, where $\mathbf{x}_t^{\mathrm{r}}$ is the pose of the end effector and τ_t^{r} the pose of the target, both in the retinotopic coordinate system.

The transformation T is updated based on the visually observed error in pointing, which is caused by the error of the model in predicting the effects of the applied action. T needs to counterbalance the visual transformation caused by the prism glasses. This can be done in different ways, but since in our experimental setup (see Section 3.1), the prism effect causes a rotation of visual observations, we chose to implement T as a rotation of the visual coordinates as well: $\mathbf{x} = T(\mathbf{x}^{\mathrm{r}}) = R^{\theta} \cdot \mathbf{x}^{\mathrm{r}}$, where the rotation

matrix R^{θ} applies a rotation over θ. The rotation angle θ is updated by the system at the end of each movement by $\theta_{t+1} = \theta_t + \eta \cdot \varepsilon$, where ϵ is the visually-observed terminal error. This error is based on the angular difference between the desired pose of the end effector at the end of the trajectory, τ_M, and the actual pose observed by the system after movement, $\mathbf{x}_{M+1}$: $\varepsilon = \angle\tau_M - \angle\mathbf{x}_{M+1}$, where M is the last time step in the action sequence. $\eta \in [0, 1]$ is the transformation learning rate, which determines the influence of the visual-shift model in the complete adaptation system.

2.3 Concurrent and terminal feedback

As in [15], we distinguish two different types of feedback: concurrent and terminal. In the first case, the end effector is continuously observed visually to obtain the internal pose $\mathbf{x}_t$, which is used by the inverse model to determine the action. This forms a visual closed-loop control system, where the effect of the action is visually observed and used in the next iteration. In the terminal-feedback condition, there is no visual closed-loop control, since the internal pose of the end effector is estimated based on proprioceptive information using the forward model. As a results, there will be a difference in terminal pointing error in the two conditions when SMCs are altered through prism exposure.

Another consequence of the feedback condition is the available data to train the inverse and forward model. In case of concurrent feedback, the new SMCs can reliably be observed over the complete trajectory. However, in the terminal condition, the SMCs related to the individual actions need to be estimated from the terminal feedback and will be incorrect due to the non-linear relations involved. We generate training data in the terminal-feedback condition by interpolating the internal visual pose of the end effector based on the visually observed end pose and the start pose estimated by the forward model.

3 Experiments

3.1 Simulation and model setup

We use a 2D simulated robotic setup to test our computational model of prism adaptation, see Figure 2a-c. The setup consists of a two degrees-of-freedom arm, a vision sensor observing the 2D position of the end effector, $\mathbf{x}_t = \{x_t^{\mathrm{x}}, x_t^{\mathrm{y}}\}$, and the target, $\tau_t = \{\tau_t^{\mathrm{x}}, \tau_t^{\mathrm{y}}\}$, and proprioceptive sensors in each of the joints, $\mathbf{p_t} = \{p_t^1, p_t^2\}$ giving information about the joint angles. In this setup, an action is a change in joint angles, $\mathbf{a}_t = \dot{\mathbf{q}}_t$. The visual observations are made from the bird's-eye perspective. In the experiment, pointing is done at two different heights, high and low, causing different arm poses during pointing. To experiment with the similarity of different poses, we add an arm-pose dimension to the input of the GPRs, with an associated length-scale parameter λ_{P}, which is by default set to 2.0 unless stated otherwise.

In a pointing trial, a target is positioned at a specific angle with respect to the robot. The system observes the target's position, $\mathbf{g}$, and then plans a target trajectory, $\{\tau_1, \ldots, \tau_M\}$, where $M = 5$ is the number of actions involved in the pointing trajectory (see Figure 2d). The prism glasses are implemented as a rotation of the visual coordinates, so that $\mathbf{y}^{\mathrm{r}} = R^{\gamma} \cdot \mathbf{y}^{\mathrm{t}}$, where $\mathbf{y}^{\mathrm{t}}$ is the true position, $\mathbf{y}^{\mathrm{r}}$ is the position in retinotopic

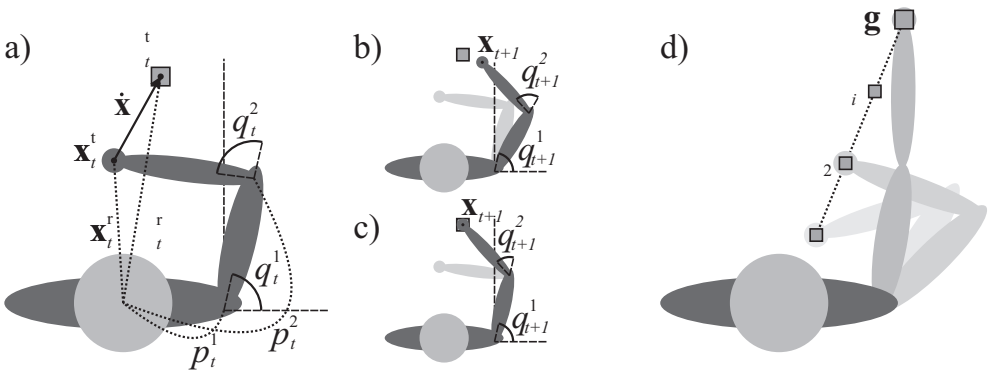


Fig. 2. Experimental setup. a) The robot has joint angles $\mathbf{q}_t$, the end effector is at true pose $\mathbf{x}_t^{\mathrm{t}}$, and the target at true pose τ_t^{t}. However, this information is not known to the system. Instead, the system observes (dotted lines) the proprioceptive signals $\mathbf{p_t}$, and the retinotopic coordinates of the end effector $\mathbf{x}_t^{\mathrm{r}}$ and the target τ_t^{r}. The model will suggest an action, which will change the joint angles. b) At first, the SMCs will not be correctly learned, and the action will result in an incorrect movement of the arm. c) After some time, the model will correctly estimate the action, so that $\mathbf{x}_{t+1} \equiv \tau_t$. d) Based on the visual target at position $\mathbf{g}$, the system plans a target trajectory giving a set of intermediate targets $\{\tau_1, \ldots, \tau_M\}$.

coordinates, and R^γ is a rotation matrix applying deviation angle γ. The application of the virtual prism glasses effects all visual observations, that is the observation of the end effector and the target.

We set the recency effect used in the forward and inverse models to $\lambda_{\mathrm{T}} = 1000$, and the transformation learning rate to $\eta = 0.35$. These parameters change the slopes of the learning curves and the influence of the forward and inverse models versus the visual-shift model. The general results are robust to small changes in these values.

3.2 Experimental setup

To compare the performance of our model to psychophysical data, we adopt the experiments performed by Redding and Wallace [15]. The task of the system is to point to visual targets. In the experiment, two different starting positions are used. In the proximal starting position, the end effector starts at the origin, and with a low arm pose. The distal starting position is halfway to the target, and with a high arm pose.

When the system has been initialized, as described below, we perform two sets of pre-tests, one for realignment and one for recalibration, to measure the performance before prism exposure. The prisms are then turned on in the simulation by rotating the visual observations over the origin with an angle of 11.4^{o}. Under prism exposure, the system performs 12 consecutive pointing movements from the distal starting position with either concurrent or terminal feedback. After that, the prisms are switched off, and the recalibration and realignment tests are performed again as post-tests. No adaptation learning is going on during the pre- and post-tests.

The realignment tests consist of a visual-shift test, a proprioceptive-shift test and a total-shift test, and are all done from the proximal starting position, i.e., different from the starting position during prism exposure. In the visual-shift test we measure subjective *straight-ahead*, by reading out $-\theta$, the negative value of the applied transformation

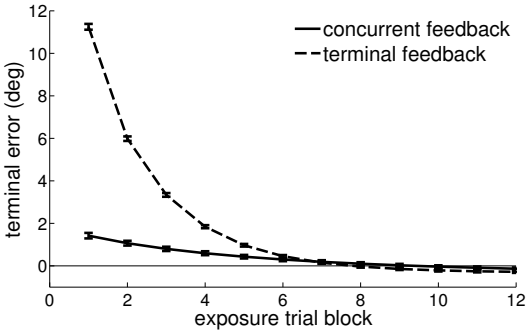


Fig. 3. The terminal pointing error as a function of pointing trial during prism exposure for concurrent and terminal feedback. The values are the means over 20 trials, and the error bars give the 95% confidence intervals.

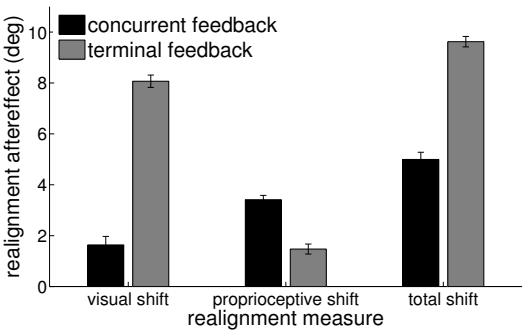


Fig. 4. The terminal error or aftereffect for the different realignment measures. With concurrent feedback, the visual shift remains small, while the proprioceptive shift is large, the reverse is true in case of terminal feedback.

in the visual-shift model. The proprioceptive-shift test measures proprioceptive straight ahead by having the system point without a visual target. To do so, the system sets a virtual target at 0° in internal coordinates. Finally, during the total-shift test, a visual target is presented at different angles $\{-20^{\circ}, -15^{\circ}, \ldots, 20^{\circ}\}$, and the system points to the targets. Pointing is always done without visual feedback.

The recalibration test is similar to the total-shift test, but the system is tested with the distal starting position, i.e., similar to the starting position during prism exposure. Moreover, we perform the test using different values for λ_{P}, to change the level of similarity in arm pose between distal and proximal starting conditions.

An experimental trial starts by initializing the system. The robot initiates 300 random movements in different parts of the work space and using the two different arm poses. Next, the system specializes on the pointing task by performing 24 pointing movements of $M = 5$ actions, using three different target positions (at -15°, 0° and 15°) and two different arm poses. Since this initialization is a noisy process and influences the performance of the system during the experiment, we repeat the experiment 20 times and report the mean values and 95% confidence intervals.

4 Results

Adaptation of pointing during prism exposure for the concurrent and terminal feedback condition is shown in Figure 3. In both conditions, the system quickly adapts to the prism effect, demonstrating accurate pointing behavior after 8 pointing trials. The error in the terminal condition is considerably larger than in the concurrent condition, which is expected, since the trajectory can be adjusted during pointing in the concurrent case. Negative error values can be observed in later pointing trials, showing overcompensation by the model, which is due to the collaboration of the forward and inverse models with the visual-shift model. These results correspond well with the psychophysical results observed by Redding and Wallace [15].

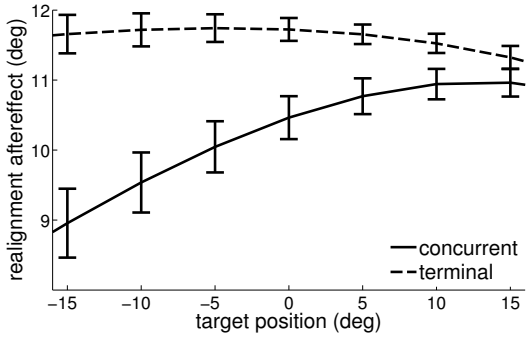


Fig. 5. The realignment aftereffect as a function of target position for concurrent and terminal feedback. The curves show different slopes depending on the feedback condition.

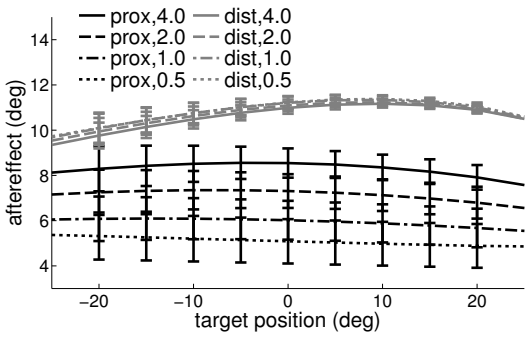


Fig. 6. The aftereffect for target positions tested at the proximal (realignment) and distal (recalibration) starting positions. λ_{P} indicates levels of pose similarity.

The size of the realignment aftereffect is shown in Figure 4. There is a clear difference in visual shift and proprioceptive shift between the concurrent and terminal condition, matching the psychophysical results [15, 14]. The visual shift is small in the concurrent condition and large in the terminal condition. In the proprioceptive test, the aftereffects are reversed, with a large effect in the concurrent and a small effect in the terminal condition. The total aftereffects are approximately the sum of the visual- and proprioceptive aftereffects.

To investigate the generalization over space, we next investigate the total-shift (realignment) aftereffects as a function of target position for concurrent and terminal feedback (see Figure 5). With concurrent feedback, the aftereffect shows a positive slope towards the prismatic shift (11.4°), whereas with terminal feedback, the curve shows a slight negative slope. This matches the psychophysical results [15].

Generalization over different actions is investigated by the aftereffects for the total shift and recalibration tests with proximal and distal starting position as shown in Figure 6. We use four different levels of similarity between the arm pose at those two starting positions, $\lambda_{\mathrm{P}} = \{4.0, 2.0, 1.0, 0.5\}$, where a higher value is a higher level of similarity. The results for the concurrent and terminal feedback condition are combined, as was done in [15]. The aftereffect is largest when the starting position is identical to the exposure phase, that is, the distal position. The curves for the distal position show a local generalization effect, with lower values for target positions that have not been used during prism exposure. This can be explained by the local learning in GPR used in forward and inverse model. The curves for the proximal position are more flat, indicating a more global generalization effect caused by the visual-shift model. Reducing the similarity in arm pose between exposure condition (distal) and the proximal condition during recalibration testing, i.e., lower values of λ_{P}, results in lower and more flat curves, indicating that aftereffects are weaker and mainly dominated by the global effect. These results correspond with psychophysical observations [15, 1].

5 Discussion

We presented a computational model capable of learning the sensorimotor contingencies involved in pointing, and of adapting to changes in these contingencies. Moreover,

it accounts for the observations made in human prism-adaptation studies [14, 15, 1], i.e. adaptation is achieved jointly through a specific local and a general global effect. Whereas local adaptation, or recalibration, is specific for the sensorimotor contingencies involved in the trained action, realignment, shows a global effect and generalizes to different actions

Using two feedback conditions, the realignment shows fundamentally different aftereffects for visual and proprioceptive tests. Where the visual shift is small and the proprioceptive shift is large for concurrent feedback, the reverse is true for terminal feedback. In our model, the visual shift is larger during terminal feedback because the observed pointing error at the end of the terminal feedback trials is larger compared to concurrent feedback trials. During concurrent feedback the end effector is continuously observed and the pointing error can be reduced while the movement is executed. In contrast, the results for the proprioceptive test are explained by the forward and inverse models. Since the sensory consequences of actions are directly observable in the concurrent condition, low-error training data is available which leads to quick adaptation of the inverse model and thus a large proprioceptive after effect. In the terminal condition, these training data need to be estimated based on the terminal observation, causing them to be less correct.

Our model gives a different explanation of prism adaptation compared to Redding and Wallace [15]. Where they consider recalibration to be a cognitive learning strategy and realignment to be an automatic process aligning different spatial maps, our model solely consists of automatic and low-level processes. Although cognitive strategies are an equally valid explanation, our model shows that similar effects can be reached with a simpler mechanism. Furthermore, in our case, the model does not include an explicit proprioceptive-shift model. Instead the results for the proprioceptive-shift test can be explained by adaptation of the forward and inverse models to changed SMCs. The presented model offers a potential system implementation for learning sensorimotor contingencies and emphasize the importance of the motor system for perception [8].

Currently, we assume a setup with a fixed eye-head system. However, in humans, both eye-head and head-hand systems are involved. The presented work is a first step towards modeling both systems. The current model is not able to learn multiple sensorimotor mappings. To account for dual adaptation effects observed in alternating prism-exposure experiments [18], future additions to the model will be necessary.

The model presented is directed at fostering our understanding of human adaptation as well as to improve on the current state of robotic systems. Inverse-kinematics learning with regression models has already been addressed by others, see e.g., [2, 9], but the addition of a recency effect and the synergy with the visual-shift model results in a fast adaptation to changes in sensorimotor contingencies.

Acknowledgments. This work was supported by the EU through the project eSMCs (FP7-IST-270212), and by the Swedish Foundation for Strategic Research.

Bibliography

[1] Baraduc, P., Wolpert, D.: Adaptation to a visuomotor shift depends on the starting posture. Journal of Neurophysiology 88(2), 973–981 (2002)

[2] D'Souza, A., Vijayakumar, S., Schaal, S.: Learning inverse kinematics. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (2001)

[3] Einhäuser, W., Martin, K., König, P.: Are switches in perception of the necker cube related to eye position? European Journal of Neuroscience 20(10), 2811–2818 (2004)

[4] Held, R., Hein, A.V.: Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. Perceptual and Motor Skills 8(3), 87–90 (1958)

[5] Held, R., Schlank, M.: Adaptation to disarranged eye-hand coördination in the distance-dimension. The American Journal of Psychology 72(4), 603–605 (1959)

[6] Kornheiser, A.: Adaptation to laterally displaced vision: A review. Psychological Bulletin 83(5), 783–816 (1976)

[7] Martin, T., Keating, J., Goodkin, H., Bastian, A., Thach, W.: Throwing while looking through prisms. ii. specificity and storage of multiple gaze-throw calibrations. Brain 119(4), 1199–1212 (1996)

[8] Nagel, S., Carl, C., Kringe, T., Märtin, R., König, P.: Beyond sensory substitution – learning the sixth sense. Journal of Neural Engineering 2, R13 (2005)

[9] Nguyen-Tuong, D., Peters, J.: Model learning for robot control: A survey. Cognitive Processing 12(4), 319–340 (2011)

[10] Nguyen-Tuong, D., Seeger, M., Peters, J.: Model learning with local gaussian process regression. Advanced Robotics 23(15), 2015–2034 (2009)

[11] O'Regan, J., Noe, A.: A sensorimotor account of vision and visual consciousness. Behavioral and Brain Sciences 24(5), 939–1031 (2001)

[12] Popović, M., Kootstra, G., Jørgensen, J.A., Kragic, D., Krüger, N.: Grasping unknown objects using an early cognitive vision system for general scene understanding. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 987–994. IEEE, San Francisco, CA, USA (2011)

[13] Rasmussen, C.E., Williams, C.: Gaussian Processes for Machine Learning. The MIT Press (2006)

[14] Redding, G.M., Wallace, B.: Components of prism adaptation in terminal and concurrent exposure: Organization of the eye-hand coordination loop. Perception and Psychophysics 44(1), 59–68 (1988)

[15] Redding, G.M., Wallace, B.: Generalization of prism adaptation. Journal of Experimental Psychology: Human Perception and Performance 32(4), 1006–1022 (2006)

[16] Redding, G.M., Wallace, B.: Intermanual transfer of prism adaptation. Journal of Motor Behavior 40(3), 246–262 (2008)

[17] Schaal, S., Atkeson, C.G., Vijayakumar, S.: Scalable techniques from nonparametric statistics for real-time robot learning. Applied Intelligence 17(1), 49–60 (2002)

[18] Welch, R., Bridgeman, B., Anand, S., Browman, K.: Alternating prism exposure causes dual adaptation and generalization to a novel displacement. Perceptual Psychophysics 54(2), 195–205 (1993)

Appendix G

Saccadic Momentum and Facilitation of Return Saccades Contribute to an Optimal Foraging Strategy

Reprinted from:

Wilming, N., Harst, S., Schmidt, N. M. and König, P. (2013). "Saccadic momentum and facilitation of return saccades contribute to an optimal foraging strategy". *PLoS Computational Biology*, 9(1), e1002871, doi:10.1371/journal.pcbi.1002871.

Saccadic Momentum and Facilitation of Return Saccades Contribute to an Optimal Foraging Strategy

Niklas Wilming[1]*, Simon Harst[1], Nico Schmidt[2], Peter König[1,3]

1 Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany, **2** Department of Informatics, University of Zürich, Zürich, Switzerland, **3** Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Abstract

The interest in saccadic IOR is funneled by the hypothesis that it serves a clear functional purpose in the selection of fixation points: the facilitation of foraging. In this study, we arrive at a different interpretation of saccadic IOR. First, we find that return saccades are performed much more often than expected from the statistical properties of saccades and saccade pairs. Second, we find that fixation durations before a saccade are modulated by the relative angle of the saccade, but return saccades show no sign of an additional temporal inhibition. Thus, we do not find temporal saccadic inhibition of return. Interestingly, we find that return locations are more salient, according to empirically measured saliency (locations that are fixated by many observers) as well as stimulus dependent saliency (defined by image features), than regular fixation locations. These results and the finding that return saccades increase the match of individual trajectories with a grand total priority map evidences the return saccades being part of a fixation selection strategy that trades off exploration and exploitation.

Citation: Wilming N, Harst S, Schmidt N, König P (2013) Saccadic Momentum and Facilitation of Return Saccades Contribute to an Optimal Foraging Strategy. PLoS Comput Biol 9(1): e1002871. doi:10.1371/journal.pcbi.1002871

Editor: Olaf Sporns, Indiana University, United States of America

Received July 26, 2012; **Accepted** November 19, 2012; **Published** January 17, 2013

Funding: This work was supported by the EU through the project eSMCs (FP7-IST-270212). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nwilming@uni-osnabrueck.de

Introduction

The effect of inhibition of return (IOR) was first described by Posner & Cohen [1]. When (covert) attention is attracted by a peripheral cue, reaction times to a subsequent probe stimulus in the same location depend in an intriguing way on the temporal offset between cue and probe: When the probe follows the cue at temporal offsets shorter than ~225 ms, fast responses are observed. In contrast, longer offsets (~225–1500 ms) lead to prolonged response times. In the original experiment, a central cross had to be fixated continuously, so the inhibitory influence at long stimulus intervals pertained to covert attention. Along similar lines, overt attention—i.e., eye movements—shows the effect of temporal IOR as well. Specifically, the fixation duration before a return saccade is on average longer compared to a saccade that continues in the same direction as the previous one. Unfortunately, several conflicting results make a comprehensive explanation of saccadic IOR and its function difficult. This study aims at a step towards an understanding of these conflicting results by further characterizing the properties of return saccades and by providing a novel view of IOR during viewing of pictures of natural and urban scenes.

But first, we shortly recap some of the discussion surrounding a functional interpretation of IOR. Posner and Cohen hypothesized that IOR might prevent the return of attention to already processed locations. A further investigation by Klein and MacInnes [2] revealed that eye movements are spatially biased away from the last (1-back) and second to last (2-back) fixation locations. This established the interpretation of saccadic IOR not only in the form of a delay, but also in spatial terms as a "foraging facilitator". That is, the function of saccadic IOR is to direct attention to unexplored parts of the stimulus, thereby fostering optimal foraging behavior. This conjecture subsequently found its way into computational models of fixation selection where saccadic IOR prevents fixating on a location twice [3–7].

Whether saccadic IOR supports such a functional "facilitator" role has been heavily discussed. There is conflicting evidence on the spatial properties of return saccades. Several studies [2,8–11] have investigated how often return saccades occur and found, depending on the precise comparison, an elevated or attenuated number of return saccades. Thus, although of crucial importance for the functional interpretation of saccadic IOR, its spatial properties are still hotly debated. There is also mixed evidence on the temporal properties of IOR. Several studies report a significantly prolonged duration of fixation before saccades to the last fixation location [2,8,12]. However, [9,13] reported a general dependency of fixation durations on the angular difference between the previous and the next saccade (termed "saccadic momentum" by Smith and Henderson). They argue that this accounts for parts of temporal IOR but that an additional localized inhibition zone remains. For saccades to the penultimate (2-back) fixation location conflicting evidence is reported whether 2-back return saccades are delayed [2,8,9,11]. In summary, the conflicting evidence of temporal and spatial properties makes it difficult to interpret saccadic IOR as a "foraging facilitator".

The dominating suggestion in the literature is that IOR supports optimal foraging strategies. This is fueled by the intuition that returning to previously fixated locations is not optimal for

Author Summary

Sometimes humans look at the same location twice. To appreciate the importance of this inconspicuous statement you have to consider that we move our eyes several billion (10^9) times during our lives and that looking at something is a necessary condition to enable conscious visual awareness. Thus, understanding why and how we move our eyes provides a window into our mental life. Here we investigate one heavily discussed aspect of human's fixation selection strategy: whether it inhibits returning to previously fixated locations. We analyze a large data set (more than 550,000 fixations from 235 subjects) and find that, returning to previously fixated locations happens much more often than expected from the statistical properties of eye-movement trajectories. Furthermore, those locations that we return to are not ordinary – they are more salient than locations that we do not return to. Thus, the inconspicuous statement that we look at the same locations twice reveals an important aspect of our strategy to select fixation points: That we trade off exploring our environment against making sure that we have fully comprehended the relevant parts of our environment.

foraging because a return saccade does not explore new parts of the environment. Hence, alternating observations of the presence/absence of inhibition of return have been taken as evidence in favor/against an optimal search strategy. However, these arguments are typically based on implicit assumptions regarding an optimal strategy and laboratory experiments with a task where it is difficult to identify the optimal foraging strategy, and therefore not based on direct investigations of fixation selection strategies. Therefore, it is presently unclear whether return fixations, contrary to the assumption that they are non-optimal, can actually be part of an optimal fixation selection strategy under natural conditions. With this in mind, we arrive at the key question of whether return locations are different from other fixation locations. For example, especially salient locations might be more likely to be fixated again, or targets of return saccades might require significantly more time to be comprehended compared to normal fixations. Such findings would suggest that return saccades might actually be due to a fixation selection strategy that needs to find a trade-off between factors such as exploration and comprehension.

We present a thorough investigation of temporal and spatial properties of return saccades by evaluating a large eye-tracking data set compiled from a host of different studies [14–17]. We analyze more than half a million fixations collected with natural scenes, urban scenes, fractals and pink noise images from 235 subjects in 5 different studies. These studies employed either free viewing conditions or a delayed patch recognition task. First, we analyze the frequency of 1- and 2-back return saccades and compare them to estimates of the number of return saccades expected from the statistical properties of single saccades and saccade-pairs. We also investigate the temporal properties of return saccades—i.e., if they are preceded by prolonged fixation durations—while paying attention to the effect of saccadic momentum. We then investigate the relationship of return locations to bottom-up saliency (as defined by local image properties). Finally, we investigate the functional role of return saccades to get a better understanding of the functional purpose of saccadic IOR and what exploration strategies could lead to the observed pattern of return saccades.

We arrive at the view that saccadic momentum can fully account for temporal IOR; that return locations are highly salient and warrant increased scrutiny by the human observer; that this scrutiny is implemented by return saccades that are observed more often than expected by chance and by increased fixation durations at return locations; and that these properties of return saccades contribute to an optimal explorative strategy.

Results

Spatial Properties of Return Saccades

We started by investigating how often return saccades occur during viewing of natural scenes. Figure 1 shows an example image with a 1-back (red) and a 2-back (blue) trajectory. Figure 2B (top left row) shows the frequency of saccade pairs with a specific amplitude and angle difference. In this plot, return saccades have a value of $\Delta\text{Angle} = 180^\circ$ and $\Delta\text{Amplitude} = 0^\circ$. We compared the number of 1-back return saccades to either the number of forward saccades or to a shuffled baseline [8] that preserved the distribution of saccade amplitudes and angles but removed order effects. The shuffled baseline accounts for return saccades due to preferences of saccade angle and amplitude combinations by the oculomotor system, but does not contain return saccades caused by facilitation or inhibition of return. In both cases, we found significantly more return saccades (bootstrapped, $p < 0.001$, Figure 3, 95% CIs created by bootstrapping per-subject percentages) in the empirical data than in the 1-back baseline. Qualitative inspection of the distribution of angle- and amplitude-differences (see Figure 3B top left row) revealed a sharp peak for return saccades ($178^\circ \le \Delta\text{Angle} \le 180^\circ$, $-1.5 \le \Delta\text{Amplitude} \le 1.5^\circ$) while forward saccades ($0^\circ \le \Delta\text{Angle} \le 2^\circ$, $-1.5^\circ \le \Delta\text{Amplitude} \le 1.5^\circ$) appeared frequently but covered a larger range of amplitude and angle differences. We also observed an asymmetry with respect to amplitude-differences. Forward saccades were often shorter than their preceding saccades (see Figure 2B top left panel). In summary, 1-back return saccades appeared much more often than expected by the distribution of saccade amplitudes and angles, and even more often than forward saccades.

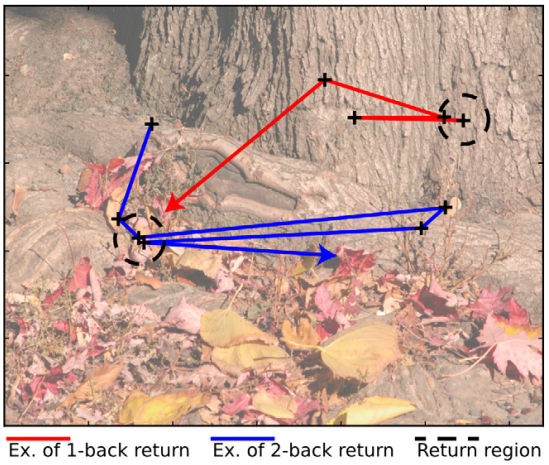


Figure 1. Example image from the category 'natural scenes'. The red line represents part of a trajectory that contains a 1-back return saccade. The blue trajectory contains a 2-back return saccade. The return region, used as a definition for return saccades for the temporal, saliency, and fixation sampling analysis is marked by the dashed circle.
doi:10.1371/journal.pcbi.1002871.g001

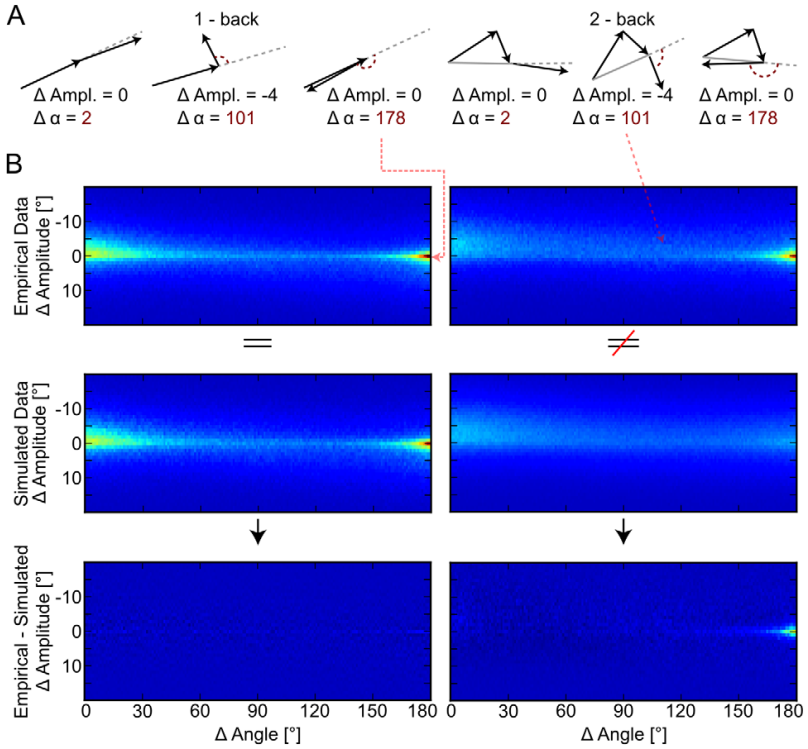


Figure 2. A shows an iconic depiction of forward ($\Delta Angle \sim 0^{\circ}$), perpendicular ($\Delta Angle \sim 90^{\circ}$), and return saccades ($\Delta Angle \sim 180^{\circ} \wedge \Delta Amplitude \sim 0^{\circ}$) in the 1 and 2-back case and their associated angle and amplitude differences. The first row in B shows the distribution of amplitude and angle differences for empirical 1-back (left) and 2-back (right) saccades. In both cases, a pronounced return saccade peak is observable. The second row shows the same, but for saccades generated with our saccade simulator. Notably, the return peak for 1-back saccades matches the peak in the empirical data while the 2-back return peak is not reproduced. The difference between empirical data and simulator output is shown in the third row (same color scheme as above). The comparison of 2-back saccades shows systematic deviations for return saccades.
doi:10.1371/journal.pcbi.1002871.g002

Next we investigated how often 2-back return saccades occur during viewing of natural scenes. While shuffling the order of saccades removes order effects for 1-back return saccades, it does not produce an adequate control distribution for 2-back return saccades. In order for this to be the case, one has to keep all 1-back return saccades due to preferences of the oculomotor system for combinations of angle and amplitudes of two consecutive saccades, but ignore all effects due to preferences of the oculomotor system for angle and amplitudes between three or more consecutive saccades (see Figure 2A, 2-back). We created control trajectories by sampling of saccades from the conditional distribution $P(L_{t+1},\Delta\alpha_{t+1}|L_t)$ (see Materials and Methods) for each subject. This distribution expresses the probability of a saccade with amplitude L_{t+1} and angle difference $\Delta\alpha_{t+1}$ given that the last saccade had amplitude L_t. It fully characterizes the angle and amplitude dependencies between two consecutive saccades but does not contain information about 2-back return saccades. To create a trajectory, we randomly drew a saccade from the distribution of first saccades for a given subject and then determined the next saccade's angle and amplitude by sampling from $P(L_{t+1},\Delta\alpha_{t+1}|L_t)$. We then iteratively added saccades to the trajectory by sampling new amplitudes and angle differences from $P(L_{t+1},\Delta\alpha_{t+1}|L_t)$, always reusing the last angle and amplitude. We matched the length of the simulated trajectories to the empirically observed lengths'.

The control trajectories reliably reproduced 1-back dependencies and the number of 1-back return saccades in particular, as well as the overall shape of the distribution of angle- and amplitude-differences between consecutive saccades (see Figure 2B, left panels). However, the control trajectories contained fewer 2-back return saccades than observed in the real data (0.0040, bootstrapped CI [0.0038, 0.0042] vs.0.0108, bootstrapped CI [0.0100, 0.0116], Figure 3). The number of 2-back return saccades was much larger than the number of forward saccades (0.0023, CI [0.00211, 0.00244], Figure 2). In fact, the 2-back histograms of the simulated and the empirical data were very similar, with the exception of the return saccade peak. We thus conclude that the statistical structure of three consecutive saccades can be explained entirely from the statistical structure between pairs of saccades, with the exception of the increased amount of return saccades.

Despite the fact that the control trajectories do not preserve statistical effects of saccade triplets and saccade quadruples, we still compared the number of 3- and 4-back return saccades to the number computed from the control trajectories. In all cases, we

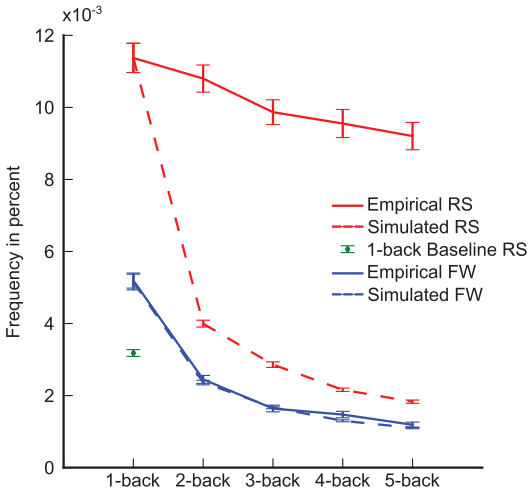


Figure 3. The frequency of return and forward saccades in empirical data and in simulation. For the case of 1-back saccades, the number of empirically observed return saccades ('Empirical RS') is larger than expected by chance ('1-back Baseline RS') and larger compared to the number of forward saccades ('Empirical FW'). The simulator reproduces the number of forward ('Simulated FW') and return saccades ('Simulated RS'). In the case of 2-back saccades, we find more return saccades than expected from the statistics of 1-back saccades, while the number of forward saccades is identical to the number of simulated forward saccades. When analyzing the presence of 3- to 5-back saccades, a similar pattern holds. Errorbars are bootstrapped 95% confidence intervals.
doi:10.1371/journal.pcbi.1002871.g003

observed many more return saccades in the empirical data (see Figure 3). We also found more return saccades than empirical forward saccades for 3- and 4-back saccades.

We conclude that locations that have been visited before are likely to be re-fixated, and for longer trajectories, this cannot be explained by the conditional dependencies between two consecutive saccades alone. We find that 1- to 4-back return saccades occur much more often than expected, but we do not observe any deviations from the predictions based on the statistics of saccade pairs for other saccades.

Temporal Properties of Return Saccades

After investigating spatial properties of return saccades we turned to temporal properties. The investigation of temporal IOR is complicated by a dependence of fixation duration on the angle and amplitude difference between the incoming and the outgoing saccade (see Figure 4B and also [9–11]). On average, it takes longer to initiate a saccade perpendicular to the last saccade relative to a forward saccade, an effect termed 'saccadic momentum'. Because this effect is reminiscent of classical IOR effects we wanted to explicitly account for saccadic momentum. To achieve this we fitted a piece-wise linear model to the fixation duration data of each subject. In a fixation sequence A→B→C the model predicted the fixation duration at location B based on the amplitude and angle differences between saccades from A→B and B→C. We used a piecewise linear model with two slopes for angle and amplitude differences respectively (Figure 4B,E). The slopes for angle differences changed at a critical angle that was fitted at the same time. However, the position of the slope change for amplitude differences was set to 0°. Please note that for visualization purposes Figure 4 shows models fitted on all data, but for the analysis models were fitted for each subject individually with a least squares procedure. The subject specific models accounted for 10% of the variance in the fixation duration data. Contrary to Smith and Henderson [9], we found that saccadic momentum did not increase linearly with the angle difference, but exhibited a change in slope for angle differences larger than 117° (CI [109, 124], slope of first segment 0.383 ms/°, CI [0.350, 0.416], and slope of second segment 0.002 ms/° CI [−0.13, 0.116], Figure 4A,B,E). The slope of the second segment is not significantly different from 0° and therefore indicates that no additional delay after the breakpoint at an angle difference of 117° occurs and that return saccades are faster than predicted by the first slope (see Figure 4B, compare red solid vs. dashed line). Thus, saccadic momentum is captured by a model with two different parts: up to angle differences of 117° fixation duration increases with 0.383 ms/° but larger angle differences do not incur a larger delay. We hypothesize that the different slopes might be due to two mechanisms that contribute to eliciting saccades with different dependencies on relative angle.

Amplitude differences changed slope at 0°, undershooting saccades had a slope of 0.39 ms/° CI [0.18, 0.60] and overshooting saccades had a slope of −2.75 ms/° CI [−3.02, −2.50].

In conclusion, the shallow slope for undershooting saccades, together with the position of the angle difference breakpoint at 117° and the 0 ms/° slope afterwards, show that the saccadic momentum effect is not specific to the return location.

Additionally we investigated IOR, similar to [9], by comparing over- and under-shooting saccades with an angle difference of 180±30°. Contrary to [9], we found no sign of a prolonging of exact return saccades (see Figure 4C, CIs for $\Delta\text{Angle} \sim 180^\circ$, $-6^\circ \leq \Delta\text{Amplitude} < 2^\circ$ largely overlap) compared to undershooting saccades. To exclude a potential effect of binning, we repeated this analysis with bins that were only one degree wide (see Figure 4D).

We also wanted to rule out the possibility that, additional to the spatially unspecific saccadic momentum effect, a spatially specific temporal IOR effect existed. We therefore fitted an 'inhibitory hill' model to the data (Figure 4F). Similar to the piecewise linear model, in a triplet of fixations A→B→C, we predicted the duration of fixation B. But this time we assumed that a Gaussian like inhibitory hill centered on fixation A would increase the duration of fixation B. The size of the inhibition was proportional to the distance between fixations A and C. We fitted this model with a least squares procedure with inhibitory hills of different sizes. The best fitting model had a Gaussian inhibitory hill with $\sigma = 3.12^\circ$ and explained 1.6% of the variance in the fixation durations. When we fitted the same model on the residuals of the piecewise linear model the variance explained dropped to less than 0.0001%. Hence, on its own the 'inhibitory hill' explains much less variance of the data than the piecewise linear fit and adding the 'inhibitory hill' model to the piecewise linear fit had virtually no benefit. We conclude that the residuals of the piecewise linear model do not contain an effect of temporal inhibition of return anymore.

Next, we investigated the effects of correcting for saccadic momentum with our piecewise linear model. Figure 4E (bottom panel and Material and Methods) shows the residuals of the piecewise linear model. We observe that fixation durations before return saccades are not systematically different from fixation durations before saccades to other locations (see Figure 4E bottom panel). In contrast, the residuals of the inhibitory hill model (Figure 4F bottom panel) show systematic dependencies on angle and amplitude differences between saccades.

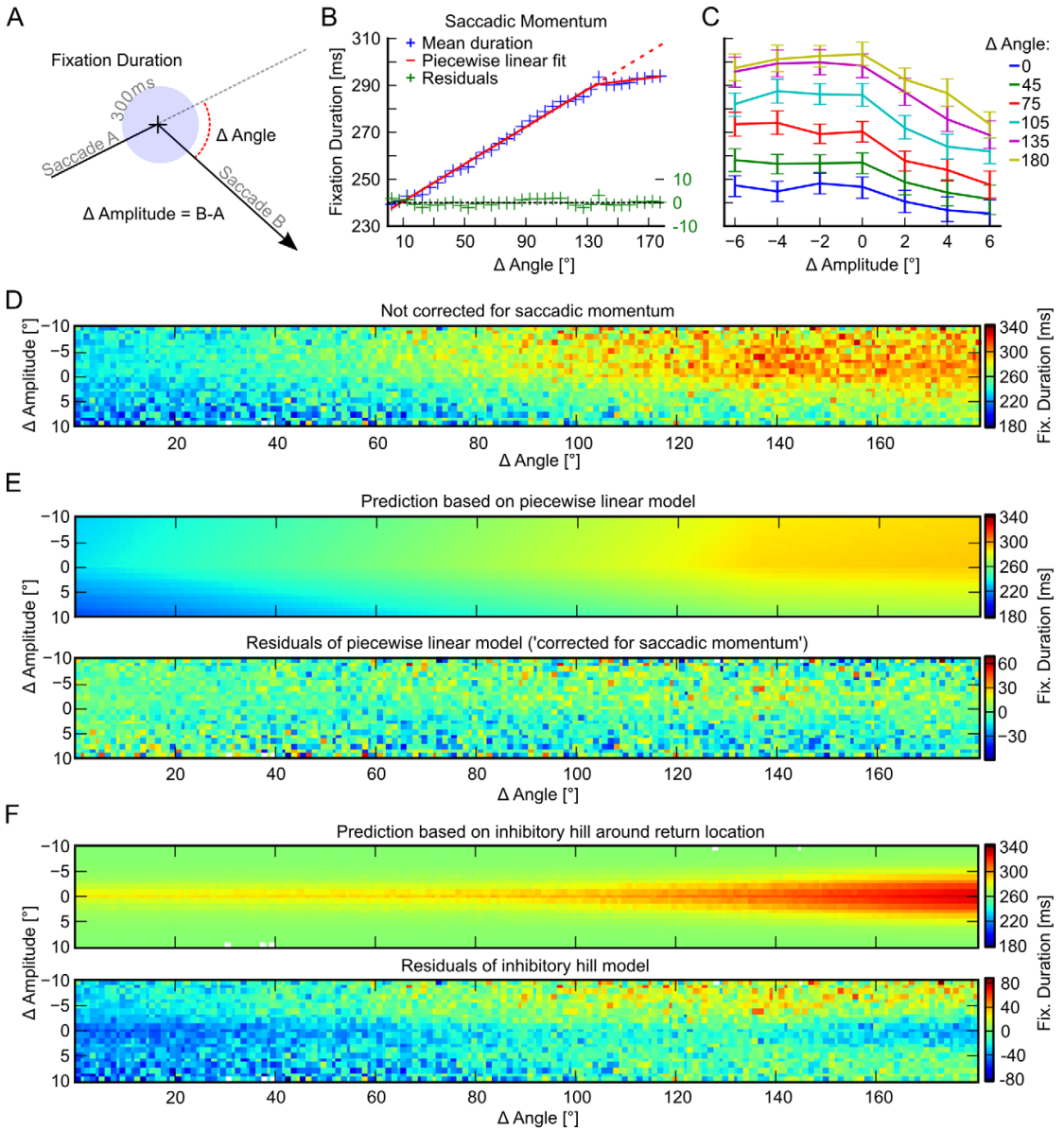


Figure 4. The saccadic momentum effect. A) Schematic drawing of plotted fixation durations, angle, and amplitude differences. B) Average fixation durations, corrected for the effect of saccade amplitude difference, as a function of the angle difference between two saccades (data is pooled over all subjects). Turning the direction of a saccade prolongs the fixation duration before the saccade is made. C) Shows average fixation durations for specific combinations of amplitude and angle differences (data is binned with bin sizes of 30° and 2° for angles and amplitudes respectively; errorbars are 95% CIs over subjects). This shows that there is no increase of fixation duration for return saccades, except for the effects of angle and amplitude differences. D) Same as C but with bin sizes of 1°; fixation durations are color-coded. E) Top panel: Prediction of average fixation duration based on the piecewise linear model (the fit is based on pooled data over all subjects for visualization purposes). Bottom panel: Residuals of correcting for angle and amplitude differences with the piecewise linear model. Here the fit was done for each subject individually, and we averaged after the correction. F) Top panel: Prediction of average fixation duration based on the inhibitory hill model (the fit is based on pooled data over all subjects for visualization purposes). Bottom panel: Residuals of the inhibitory hill model. Here the fit was done for each subject individually, and we averaged after the correction.

doi:10.1371/journal.pcbi.1002871.g004

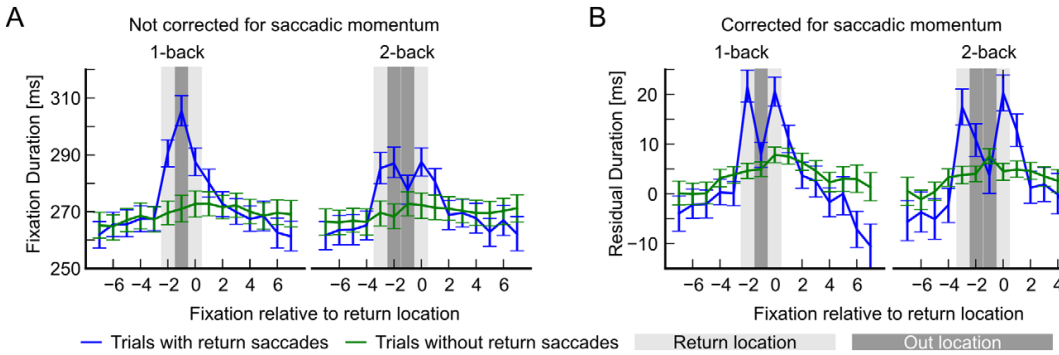


Figure 5. Fixation durations at return locations are longer. A) The average fixation duration at return locations in trials with 1-back and 2-back return saccades (blue lines) is longer than fixation durations in control trajectories (green lines). Errorbars are 95% CIs bootstrapped over subjects. B) Correcting for saccadic momentum with the piecewise linear model completely removes any trace of temporal inhibition of return for 1- and 2-back return saccades. Errorbars are 95% CIs bootstrapped over subjects.
doi:10.1371/journal.pcbi.1002871.g005

In summary, the prolonging of fixation durations before return saccades can be explained in terms of saccadic momentum and saccadic momentum is not specific to return locations.

We next considered fixation durations at return locations to investigate if they are looked at more often because they were not scrutinized sufficiently the first time around [8] or because they are highly salient and also demand above-average processing time.

To this end, we compared all trials (i.e. the entire fixation trajectory of one subject on one image) that contained return saccades (RS-trials) with all trials that contained no return saccade. We centered all RS-trials on the 2nd fixation of the return location. We aligned trials of the same length without RS to the trials that contained a RS. If for example, the 2nd fixation of the return location occurred at fixation Nr. 5, both trials were centered on fixation Nr. 5. Figure 5A shows that fixation durations at the return location are significantly longer than at control locations. Remarkably, this even holds when the location is visited for the first time. We observed the same pattern with a reduced effect size for 2-back return saccades. Hence, return saccades are not due to a shortened analysis at first fixation; fixation duration is significantly increased during first fixation and re-fixation.

Please note that correcting for saccadic momentum and saccade amplitude differences eliminates the increase in average fixation duration before the return movement, where IOR has been typically observed (see Figure 5B). This supports our conclusion that controlling for the effects of saccadic momentum explains the prolonging of fixation durations before return saccades in our data.

To check if saccadic IOR effects that could not be explained by saccadic momentum were present in the individual experiments that we analyzed, we repeated the comparison of RS-trials and non-RS trials for every dataset. We checked if the difference at the out-location between both trial types was significantly different from zero when we corrected for saccadic momentum with our piecewise-linear model. We did not find any significant deviations (paired T-test, $p>0.05$, Bonferroni corrected).

In summary, in our data temporal effects of IOR can be accounted for by a pronounced, non-linear effect of saccadic momentum and saccade amplitude differences, which is not specific to the return location. Additionally, the average fixation duration at the return location is longer for 1- and 2-back saccades, already during the first visit.

Return Saccades and Saliency

The observation of increased fixation duration at return locations suggests that such locations are special. To investigate whether the stimulus was systematically different at return locations compared to regular fixations, we computed bottom-up saliency at both locations based on the values of a large number of low (e.g. luminance, red-green and blue-yellow contrast) and mid-level (e.g. symmetry, intrinsic dimensionality) stimulus features.

We compared the values of 63 local features (please see Materials and Methods: Feature Analysis for the complete list) at return and non-return (normal) fixations in the dataset used in [17] (Figure 6A,B). For quantification, we computed the area under the receiver-operating characteristics curve (AUC) of a linear classifier that separates return and normal fixation locations from control locations on the same image (Figure 6C) [18]. The AUC measures how well a feature can be used for correct classification. 0.0 implies perfect classification but switched labels; 0.5 is chance performance; 1.0 is perfect. Control locations were sampled from all fixation locations made on other images from the one in question and hence take into account the general spatial bias. We calculated the AUC for separating return fixation locations from controls and the AUC for separating normal fixation locations from controls.

We observe a linear relationship between AUCs of different features calculated for return-locations and normal-fixations. Furthermore, this holds for natural and urban scenes (Figure 6D, each data point shows AUC values for one image feature). The pattern of AUC values for return and normal fixations is well described by a linear relationship (natural scenes: $r^2=0.76$, urban scenes: $r^2=0.95$). Only the phase congruency feature does not fit this linear pattern; it is slightly better for predicting normal fixations than return fixations (lower left corner in left panel of Figure 6 D). Importantly, the slope of the linear fit is less than 1.0 (natural scenes: $\beta=0.56$, T-test $\beta<1 : p<0.0001$; urban scenes: $\beta=0.77$, T-test $\beta<1 : p<0.0001$). Hence, those features that predict normal fixation locations above chance (AUC>.5) better predict return locations than regular fixation locations. Importantly, those features that are anti-predictive (AUC<0.5) are also more anti-predictive of return locations than of regular fixation locations. This indicates that the pattern of contribution of different features, as quantified by the AUC values, does not differ between normal and return locations. Such a linear relationship implies that image feature based salience models trained only on

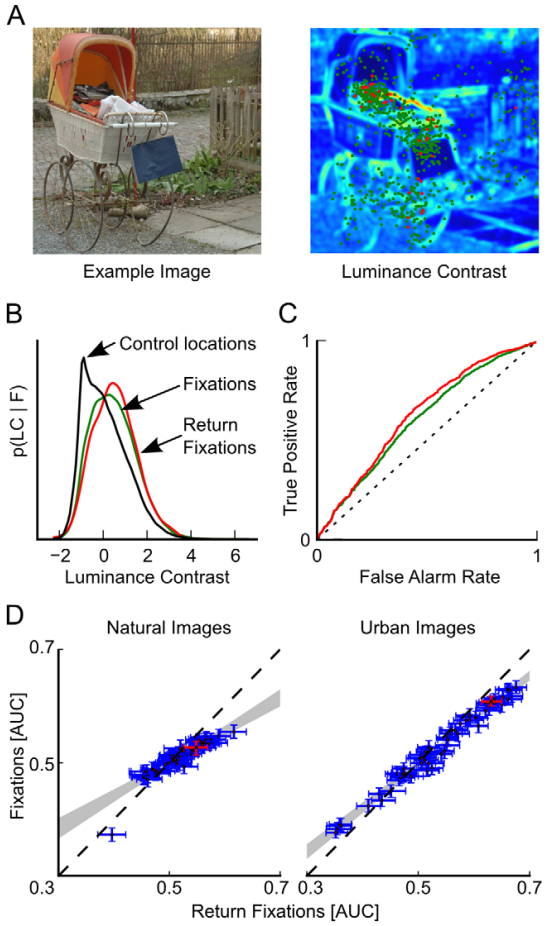


Figure 6. Image features predict return locations better than normal fixation locations. A) The right panel shows the luminance contrast feature for the image on the left. Green dots mark regular fixations, and red dots mark return locations. B) The distribution of feature values at control locations, regular fixation, and return fixation locations. C) The ROC curve for separating regular and normal fixations from control locations. D) AUC values of individual image features for return and regular fixation locations. Return locations are systematically better predicted by image features than regular fixations—i.e., return location feature AUCs are higher for predictive features (AUC>.5) and smaller for anti-predictive features (AUC<.5). Error bars are bootstrapped 95% CIs. The relationship between regular feature-fixation AUCs and return feature-fixation AUCs is well described by a linear relationship ($\langle r^2 \rangle = 0.85$, $\langle \beta \rangle = 0.66$). Gray shaded area: convex hull of regression fits between return and regular feature AUC patterns.
doi:10.1371/journal.pcbi.1002871.g006

regular fixation locations will perform better on return locations than on regular locations. In summary, image features better predict return locations than regular locations.

To compare bottom-up saliency values at return and normal fixation locations, we compiled a weighted sum of all 63 features into a single saliency score. Weights for the linear combination were obtained by a logistic regression that separated either return locations from controls (RS-model) or normal fixations from controls (FIX-model, see Materials and Methods). We then computed the AUC of both saliency scores for separating

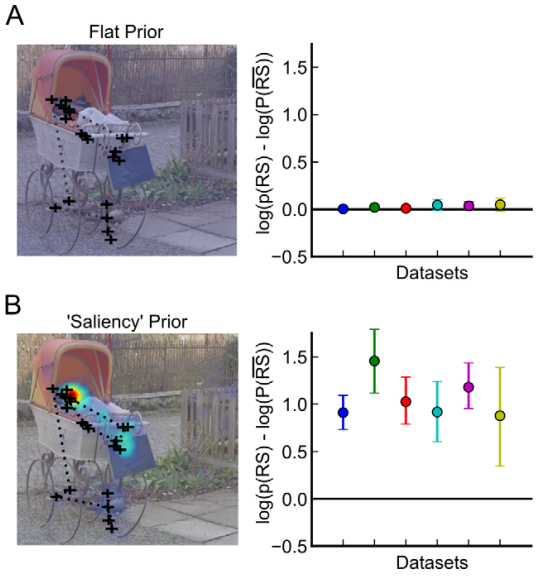


Figure 7. Return saccades increase the trajectory likelihood when observers sample from an empirical saliency distribution. Both plots show the difference of the log likelihood for trajectories with and without return saccades as a function of the dataset. A) If all locations have equal probability of fixation, trajectories with return saccades are as probable as trajectories without return saccades. B) If salient locations are more probable than other locations, trajectories with return saccades are more likely than others. Error bars are bootstrapped 95% CIs.
doi:10.1371/journal.pcbi.1002871.g007

return-locations and non-return locations from controls. We used leave-one-subject-out cross validation to ensure independence of training and test data. We found that return locations could be better predicted (average AUC of 0.733; RS-model: 0.731, FIX-model: 0.736) compared to normal fixations (average AUC of 0.670; RS-model: 0.667, FIX-model: 0.674). From this analysis, we conclude that return saccades are directed to more salient locations than normal saccades and that the pattern of feature-fixation correlations is comparable to return locations and normal fixations.

Fixation Sampling Strategies

The finding of an increased number of return saccades and prolonged fixation durations at return locations is difficult to reconcile with a foraging strategy that maximizes the entropy of a fixation density map, i.e. the area that is 'covered' by fixations. Yet return locations are 'special' in the sense that they are looked at longer and do not appear at random locations. Instead of maximizing entropy, we hypothesize that the very existence of return fixations serves to optimize the match of saccadic trajectories with an internal priority map that encodes which locations are relevant in the scene. Here we replace the spatially flat prior of the maximal entropy assumption (Figure 7A) with a stimulus-dependent prior and use the viewing behavior of (other) subjects as a proxy for such an internal priority map (Figure 7B). That is, we use empirically defined salience, given by how often different subjects look at a location, as the internal priority map.

In this respect, we were interested if, all else being equal, a return saccade would increase the probability of a trajectory according to the internal priority map. We compared trajectories

with return saccades to the same trajectories that, instead of exploiting an already seen location, explored one additional new location.

More specifically, for each fixation trajectory that contained a return saccade, we first computed a fixation density map from the fixations of all other subjects on the same image. We made sure that in this computation, trials containing return saccades were omitted (see Materials and Methods). We then used this fixation density map as an internal priority map for the trial in question. We compared the probabilities of generating two different kinds of trajectories based on the fixation trajectory in question from this internal priority map: The first contained the return saccade (return-trajectory) but we removed the last fixation. For the second trajectory (exploration-trajectory) we removed the 2nd visit to the return location but kept the last fixation. The exploration and return trajectories thus contained the same number of fixations, but the exploration trajectory contained one more unique fixation location (see also Materials and Methods). In other words, given the original fixation trajectory A-B-A-C … -F-G, the return trajectory is given by A-B-A-C…F and the exploration-trajectory is A-B-C…-F-G.

The probability for the exploration and return-trajectories was defined as the probability to draw exactly these trajectories from a multinomial distribution with event probabilities given by the internal priority map. Because we use a multinomial distribution as our model, the order of fixations is irrelevant and changed distances between fixations do not confound the results. We find that return saccades actually increase the probability of a trajectory compared to the omission of such saccades (Figure 7, ANOVA with factors experiment and saliency map type, main effect of saliency map type $p < 0.0001$, no other significant effects at $p = 0.05$).

In summary, to match an internal priority map it is better to allow return saccades to exploit empirically salient locations in the priority map compared to forcing all saccades to unexplored locations. This result is also reflected in the additional finding that return locations show higher average values of the internal priority map compared to locations before and after return locations. That is, humans try to visit empirically highly salient regions but trade off exploitation and exploration by revisiting important parts of the stimulus.

Discussion

In this study, we investigated the spatial, temporal and functional properties of saccadic inhibition of return.

With respect to spatial properties, we find more 1-back and 2-back return saccades than expected from the distribution of saccade angles and amplitudes and relative angles and amplitudes. Also, our novel statistical model for 2-back return saccades reproduces the distribution of angle and amplitude differences of saccade triplets very well except for 2-back return saccades. This indicates that our model is adequate to explain higher order biases in saccade trajectories but that 2-back return saccades are facilitated compared to these higher order biases.

This agrees with findings from Hooge, Over, van Wezel and Frens [8], who used a comparable baseline for 1-back return saccades. Smith & Henderson [9–11] used two different baselines but find similar results. Compared to distance matched controls (e.g. saccades with $\Delta\text{Amplitude} = 0^\circ$ and $\Delta\text{Angle} = 90^\circ$) they report an equal or larger number of 1-back and 2-back return saccades. Compared to a baseline where the order of fixations is shuffled they report more 1-back and 2-back return saccades in their empirical data.

In disagreement with our results Bays and Husain [19] argue that 1-back return fixations occur less often than should be expected. The critical difference to our study is the baseline used for comparing the number of return saccades. Bays & Husain argue that saccade trajectories are not only influenced by oculomotor biases but also by the spatial distribution of salient locations in an image. They generate control trajectories that take both biases into account by sampling from the conditional probability distribution $P(x_t = X | x_{t-1} = Y)$, which expresses the probability to fixate location X given that the current fixation is at location Y. Importantly, the resulting trajectories contain more return saccades than their empirical data. Because the process that generated these trajectories did not take into account past fixations but still created more return fixations, Bays and Husain conclude return locations are actively inhibited. What could explain the differences between our and Bays & Husain's findings? There are several differences regarding the acquisition and analysis of eye-tracking data in our and Bays and Husain's study. First, Bays and Husain presented images for 20 s while in the present investigation the presentation time was 6 s or shorter. Fixation trajectories over repeated presentations of the same stimulus are partly overlapping [16], supporting the argument that prolonged stimulus presentations might have an effect on the frequency of return-saccades. Second, we seem to observe a much more localized return peak (compare our Figure 2B and their Figure 1B). This percolates to differences in the definition of return saccades between Bays & Husain's and our work and thereby to a different estimate of the number of empirical return saccades. Third, to accurately estimate conditional probability densities a large amount of data is required. In the present study we opted for a reanalysis of several previously conducted studies resulting in a very large database. This allowed us to remove trajectories containing return saccades from the estimate of fixation probability densities. Fourth, typical laboratory setups with limited size monitors enforce saccades with larger than 90° turning angle in order to maintain the gaze within the monitor boundaries, part of these are classified as return saccades. To resolve these issues and reach a final conclusion more research is warranted.

With respect to the temporal properties of return saccades, we find that direct return saccades are preceded by longer fixations than forward or perpendicular saccades. We therefore replicate a classical effect of saccadic inhibition of return. However, this effect is explained by saccadic momentum [9,13]: A piecewise-linear dependence of fixation durations on angle and amplitude differences between the preceding and next saccade, i.e. large turns in eye movement direction are preceded by longer fixation durations compared to small turns. We find that exact return saccades do not take longer than undershooting return saccades, or saccades with an angle difference larger than ~117°. Correcting for the effect of saccadic momentum removes the delay imposed on direct return saccades compared to non-return saccades. Crucially, we find that the dependence of fixation durations is constant after a critical angle. That is, return saccades are faster than expected from the slope of saccadic momentum before the critical angle. We furthermore tested directly if a localized 'inhibitory hill' around the return location could explain our data and found that it performed worse than the piecewise linear model. Additionally, the inhibitory hill model did not explain variance in our data that was not explained by the piecewise linear model. In conclusion, we did not find a spatially localized inhibitory effect for return saccades in addition to saccadic momentum. We therefore conclude that our data is better described by a spatial facilitation of return, a delay increasing

linearly with angle of changing gaze direction (saccadic momentum) up to a critical angle and constant thereafter.

Our results are compatible with many findings in the literature. We replicate the classical saccadic IOR effect [2,8] on a large data set and provide a detailed description of saccadic momentum [9–11,13]. However, we could not replicate Smith & Henderson's [9] finding of an extra delay, in addition to saccadic momentum, for return saccades. Also, our results contrast with findings of Hooge and Frens [12] who find a localized temporal zone of inhibition of 4° by asking their subjects to carry out a predetermined sequence of saccades. The difference between Hooge et al.'s and our results might be explained by the two very different tasks and stimulus arrangements. It is well known that the oculomotor system in the brain includes many different areas that are tightly coupled [20]. Carrying out pre-programmed saccade sequences might recruit neural substrate that elicits temporal inhibition of return. Hooge et al. suggest that the superior colliculus might be the neural substrate that causes effects observed in pre-planned saccade sequences. In contrast, free viewing, where fixation locations are selected based on local salience, oculomotor bias's and other top-down factors, might activate different networks that lead to different temporal properties of fixations. One candidate would be LIP which has been implicated in computing a priority map [21] which combines bottom-up and top-down information to guide selection of fixation targets during visual search.

An alternative non-exclusive explanation, that would incorporate both contradicting results, might be that precise saccadic IOR can be tuned by the visual system. This is supported by a study from Farell, Ludwig, Ellis, and Gilchrist [22] that shows that the classical IOR effect is adaptive to environmental statistics. However, because they did not explicitly investigate saccadic momentum, it remains to be seen what is modulated: a return location unspecific saccadic momentum, return location specific IOR or both.

Interestingly we find that return locations are more salient, according to empirically measured as well as stimulus dependent saliency, than regular fixations. Hooge et al. [8] find that the first visit of a return location is shorter than the second visit. They suggest that return saccades occur because the visual system did not have enough time to analyze a fixation location during the first visit. In our data return locations are fixated longer compared to control fixation locations during both visits. The visual system therefore has more processing time available for return locations than for regular locations. These findings suggest that return locations need to be scrutinized in more detail than regular fixation locations.

We also found that return saccades increase the match of individual trajectories with a grand total priority map. The priority map was defined by empirical salience, i.e. those locations that are consistently fixated by many subjects. Because trajectories that contained return saccades were more likely than trajectories that explored a new location with every fixation, we suggest return saccades are the consequence of a fixation selection strategy that samples relevant parts of a scene. Furthermore, because the internal priority map was defined by empirical salience, which we interpret as a proxy for behavioral relevance, return locations were more relevant than other locations. We therefore suggest that the fixation selection strategy trades off exploration of unseen relevant locations and exploitation of already seen relevant locations with return saccades.

What Are Possible Mechanisms that Could Explain Our Findings?

With respect to saccadic momentum, the question arises whether the observed regularities could be an effect of the physical properties of eye-movement control. Different patterns of muscle movements are necessary for return saccades and forward saccades. Forward saccades require flexed muscles to be flexed more, while stretched muscles must be stretched more. Return saccades require an inversion of these muscle states: flexed muscles must be stretched and stretched muscles must be flexed. This might contribute to the observed differences in fixation durations. However, when talking about muscle effects, two things should be kept in mind: First, the temporal difference between the length of fixations before return and forward saccades is in the order of 50 ms (Figure 4B). Bahill, Clark and Stark [23] show that saccades of up to 20° can be completed in less than 60 ms and thus it is safe to assume that the time needed for the acceleration of the eye during a saccade is much shorter than 50 ms. Therefore, differences in activation patterns on the muscular level cannot explain the systematic increase of fixation durations observed here.

Second, Farell, Ludwig, Ellis, and Gilchrist [22] found that the temporal difference between return and forward saccades is modulated by the likelihood of a return saccade, with the effect eventually vanishing when return saccades are very likely. While this does not rule out a contribution of muscle effects to saccadic momentum, the least it demonstrates is that saccadic momentum can be modulated by factors that are independent of physical motor control.

Ludwig, Farell, Ellis, and Gilchrist [24] proposed that 'Inhibition of Return' can be explained in a decision-making framework. In short, potential saccade targets accumulate evidence until an evidence threshold is reached. The first target that reaches the threshold is used as the next saccade target. Indeed, they find a difference in saccade latency for return and non-return saccades that correspond to differences of accumulation rate in their fitted models. However, they only differentiate return and non-return saccades. Thus, the phenomenon of saccadic momentum makes a large contribution to that comparison and might easily dominate. It would be interesting to see if the accumulation rate is parametrically modified by the angle difference between the new saccade target and the last saccade. However, even if a change in accumulation rate can explain saccadic momentum, this would make the high incidence of return saccades even more puzzling. Furthermore, the question remains why the accumulation rate changes with changes in eye-movement direction.

But we find an alternative suggestion more tenable: If we imagine that we shift the center of gaze from point A to B, then parts of the stimulus between A and B will have been sampled by the fixation of A. Thus, relative to B, backward targets are at locations for which prior information exists while forward targets deal with parts of the stimulus for which no (or less) information is available at that moment in time. We hypothesize that forward and backward targets have different accumulation rates because different amounts of knowledge are available for these locations. Considering that receptive fields are remapped during saccades, it does not seem unlikely that prior knowledge is transferred during the remapping [25]. For salient targets to reach the decision threshold faster when they are 'forward' compared to 'backwards', the accumulation rate has to be inversely proportional to the amount of knowledge available. This would imply that more prior knowledge leads to slower accumulation of evidence. It seems that such a notion is compatible with accounts of predictive coding in which higher-level information explains away activity at lower levels [26]. Here the higher-level knowledge about backwards locations 'explains away' their salient properties, thereby making them less salient compared to forward locations. This in turn would lead to a slower accumulation of evidence at backward compared to forward locations. In that context, the observed

piecewise linear dependence of fixation duration on saccadic angle is important. Based on this observation we suggest that two competing mechanisms are active in parallel, and only one of these—the one with a steep dependence of fixation duration on the saccadic angle—is dependent on the already available knowledge. While this is speculative, the proposal does fit with our finding of increased bottom up saliency at return locations.

In summary, accumulator models are a promising tool to understand the dynamics of saccade target selection. Future studies will need to link saccadic momentum and facilitation of return to specific properties of such models. The findings that return locations are more salient and looked at longer must be crucial parts of this puzzle.

What Could Be the Function of Facilitation of Return and Delay of Direction Change?

Clearly, spatial facilitation of return is incompatible with the objective of covering the entire stimulus evenly with fixations in a short amount of time. However, what is the motivation to assume that the stimulus is equally interesting in all locations? In an everyday search task, such as when looking for the car keys, one would not cover all places from cellar to rooftop evenly. Instead, it is sensible to scrutinize those locations that are likely due to prior knowledge and to look twice before considering more exotic alternatives. Under laboratory conditions, for example when a near threshold Gabor patch is superimposed on a pink noise image at a random location, the search strategy might adapt to the flat location prior [27]. This is a remarkable feat of behavioral adaptation, yet no reason to assume that return saccades are generally inhibited. During free viewing, no explicit external task is enforced and subjects do not relate their eye-movement behavior to an externally set optimality criterion. Some studies included in our data set employ specific tasks. Specifically, in the delayed patch recognition task [14], subjects have to decide whether a target patch was contained in a previously shown sample image. The target patches are selected uniformly from the entire stimulus, which might suggest that return saccades are not useful to solve the task. However, the probe patch is not presented in the location where it was in the stimulus and after stimulus offset only. This makes keeping track of where in the sample image the target patch was selected difficult. Furthermore, to prevent fatigue, [14] deliberately choose to present only 128 images to each subject. The number of trials is therefore considerably smaller than in psychophysical studies with reduced setups. Hence, the opportunity for subjects to infer and adapt to the objective prior of patch locations is rather limited. But even for fully adapted subjects, it is unclear whether seeing the entire stimulus is the optimal strategy for a delayed patch recognition task. The task requires not only passive observation of the stimulus but encoding and recalling as much as possible of it at a later stage. The optimal strategy needs to trade off holding complex stimulus patches in memory with exploring new parts of the stimulus. In this respect, return saccades might be part of an optimal strategy because they allow the visual system to exploit information at relevant locations more thoroughly.

It could be argued that we did not use a visual search task and therefore found more return saccades than expected. As described above two studies included in our data set employed a delayed template match search task where homogeneously distributed fixation locations seem advantageous. Furthermore, even during visual search return saccades are not automatically disadvantageous for search performance. First, a consistent central bias has been documented in many studies (for example [17,28]), invalidating an assumption of a flat prior. This shows that the visual system does not consider every location to be equally relevant. Second, Hooge, Over, van Wezel and Frens [8] find more 1-back return saccades than expected during a visual search task. Third, even during visual search, a single fixation might not suffice to identify a target in front of the background, and there is evidence that humans take uncertainty inherent in their visual system into account [27]. Also, there clearly are prior expectations about where targets of specific types can be found in a scene [29] (e.g., pedestrians are usually not located in the sky). These two conditions necessitate trade off of exploration and exploitation in visual search—return saccades (exploitation) with saccades that target unseen parts of the stimulus (exploration). Therefore return saccades are likely to be a part of visual search strategies as well.

Having considered everyday search tasks, free viewing, and delayed patch recognition, we find it unconvincing that a flat spatial prior over stimulus locations must be part of a good strategy to solve these tasks. In turn, we argue that from the existence of return saccades, it does not follow that a task is not being solved optimally.

Instead, a novel view concerning the functional interpretation of IOR emerges. Farell et al. [22] have shown that the time difference between return and forward saccades is adaptive to the environment. Smith & Henderson [10,11] argue that saccade latencies are the result of several interacting processes such as bottom-up input, top-down control and saccadic momentum. We provide evidence for the hypothesis that return saccades are part of a strategy that aims at devoting attention to the most relevant information in the stimulus: First, return locations are more bottom-up salient than regular fixation locations, showing that the stimulus is different at return locations compared to regular locations. Second, return locations are fixated longer during both visits, indicating that more attention compared to regular fixation locations is devoted to return locations. Third, return saccades occur more often than expected, suggesting that they are an important part of a fixation selection strategy. Most importantly, if we accept eye movement behavior of other subjects as a proxy for relevance, return saccades increase the likelihood of a trajectory to sample the relevant parts of an image. We therefore conclude that spatial facilitation of return, saccadic momentum, and relative speed-up of saccades at very large angle differences might not serve a single objective but might emerge from the broader goal to optimally sample relevant parts of a stimulus.

Materials and Methods

Data

We re-analyzed data from several studies conducted at the Institute of Cognitive Science, University of Osnabrück. Here we briefly summarize the different studies but leave details to the respective original publications. Açik et al. [14] investigated the effect of age on viewing behavior. They presented 128 images from the categories 'manmade scenes', 'natural scenes', 'fractals', and 'pink noise' for 5 s. The images were selected from a larger database that contained 64 images per category. Images from the same database were used in [15,17]. After each image, an image patch was presented, and subjects had to answer whether this patch was contained in the previously shown image. Fifty-eight subjects participated in this study (18 elementary school children with a mean age 7.6, 23 university students with a mean age of 22.1, and 17 older adults with a mean age of 80.5). Wilming et al. [17] showed 128 images from the categories 'manmade scenes' and 'natural scenes' in a free viewing paradigm with a viewing duration of 6 s to 48 subjects (aged 19 to 28 years). Kaspar and König [15] investigated the influence of repeated stimulus

presentations, image category, and individual motivations. They presented 48 images taken from the same scene types used by [14] and repeated the presentation of each image 5 times. The subjects were instructed to freely view the image for a period of 6 s. Forty-five subjects participated in the study (aged 18–48 with a mean age of 24.2 years). Kaspar and König [16] (data from 'Experiment 2') presented 30 different urban scenes to 34 subjects (aged 19–49, mean age 25.9 years) with a viewing duration of 6 s. Each image was presented five times to each subject. The images were not part of any of the other studies used here. We analyzed data from two more experiments; the results have so far not been published. In these studies, conducted by Alper Açik, 50 subjects were presented with contrast modified and phase scrambled images from the category 'fractals'. After a stimulus presentation of 5 s, subjects had to perform a 2AFC patch recognition task (20 subjects) or a YES/NO patch recognition task (30 subjects). We treated the two different tasks as different datasets.

All studies used an Eyelink II eye-tracker (SR Research Ltd., Mississauga, Ontario, Canada). All studies were conducted in compliance with the Declaration of Helsinki as well as national and institutional guidelines for experiments with human subjects. Because different studies used different displays and image sizes, we converted all fixation coordinates into degrees of visual angle. In total we analyzed over 597,000 fixations collected from 235 subjects in 6 different studies.

Spatial Properties of Trajectories

To investigate the frequency of 1- and 2-back return saccades, we created two different baseline conditions. For the 1-back condition, we shuffled all of the recorded saccades. This removed all order effects but did not change the distribution of saccade angles and amplitudes. We used this shuffled baseline to estimate how many return saccades should be expected by randomly sampling from the distribution of saccade angles and amplitudes. All saccades with an angle difference larger than 178° and amplitude difference of less than $\pm 2^\circ$ were considered return saccades. To determine significant deviations of the number of return saccades from the shuffled baseline, we bootstrapped 95% confidence intervals around the mean difference of return saccades for each subject and checked if the confidence interval contained 0. In comparison, the empirical data contained significantly more return saccades in the 1-back condition. Bootstrapping the per-subject percentages created the 95% confidence intervals shown in Figure 3.

Subsequently, to investigate whether 2-back and higher dependencies between saccades can be explained by 1-back information, we devised a saccade generator, which uses 1-back information of trajectories as an input to generate arbitrarily long sequences of saccades. As the generator does not use any 2-back information, any patterns that can be observed in the 2-back condition of the generated data are due to 1-back dependencies alone. The generator creates a trajectory by drawing a saccade from the distribution of first saccades in the input data and copies its absolute angle and amplitude. Subsequently, further saccades are added by drawing their angle difference and amplitude with respect to the last saccade from the conditional distribution $P(L_{t+1},\Delta\alpha_{t+1}|L_t)$. This distribution expresses the probability of observing an amplitude difference L_{t+1} and angle difference $\Delta\alpha_{t+1}$ at the next saccade, given that the length of the last saccades was L_t. It thus comprises only 1-back information. We estimated this distribution for every subject separately by computing histograms for each possible value of L_t. Sampling from the non-conditional probability distribution from Figure 2 does not generate valid adjoining saccade trajectories because not all negative amplitude differences can be generated at all times. In terms of fixation coordinates, no additional restrictions were made such that the simulator precisely replicates 1-back dependencies without incorporating any additional image statistics such as picture size. The resulting set of fixations could be analyzed in terms of saccade dependencies equal to the empirical data. To validate the accuracy of the saccade simulator, we compared the similarity between subjects and the similarity between subjects and simulator. To this end, we computed for each subject the distribution $P_{emp}(\Delta L,\Delta\alpha)$ of amplitude differences ΔL and angle differences $\Delta\alpha$ (see Figure 2) for 1-back saccades. Subsequently we computed $P_{sim}(\Delta L,\Delta\alpha)$ for each subject based on saccades generated from their own distribution $P_{emp}(\Delta L,\Delta\alpha)$. Finally we computed the KL-divergence between subjects and between subjects and their simulated saccades:

$$\left\langle D_{KL}\left(P_{emp}(\Delta L,\Delta\alpha)^i \middle\| P_{emp}(\Delta L,\Delta\alpha)^j\right)\right\rangle \forall i,j \wedge i\neq j$$

$$\left\langle D_{KL}\left(P_{emp}(\Delta L,\Delta\alpha)^i \middle\| P_{sim}(\Delta L,\Delta\alpha)^i\right)\right\rangle \forall i$$

where i and j are subject indices. We found that the KL-divergence between subjects was higher than the divergence between subjects and simulator output. Additionally, the number of return saccades generated by the simulator is not different from the number of return saccades found in the empirical data. Furthermore, qualitative comparison of differences between empirical and simulated data did not reveal any systematic deviations in the 1-back case. From this, we conclude that the simulator reliably replicates all 1-back dependencies in the data.

To compare the number of return saccades, we again bootstrapped 95% confidence intervals around the difference of simulated and empirical return saccades and checked if the interval contained 0. As expected, this was the case for 1-back saccades. All other comparisons showed significantly more empirical return saccades (see Figure 3). In the case of forward saccades, all comparisons contained 0.

To assess the similarity of the distributions $P_{emp}(\Delta L,\Delta\alpha)$ and $P_{sim}(\Delta L,\Delta\alpha)$ for the 2-back case (see Figure 2, left column), we calculated the KL-divergence between the two for each subject. The mean KL-divergence was 0.21, to which the return peak contributed more than any other area of comparable size (for example, 4 times as much as the forward peak). Thus, all other 2-back dependencies were very similar.

Temporal Properties of Return Saccades

Because effects of saccadic momentum on fixation durations are largest at return locations, they potentially confound findings of IOR. [9,11] considered the effect of saccadic momentum by comparing average fixation durations for exact return saccades and over- and under-shooting return saccades. We repeated this analysis but take several other measures to ensure a fair comparison. First, we explicitly estimated the effect of saccadic momentum and saccade amplitude differences on fixation duration with a non-linear breakpoint regression:

$$y=\beta_1\Delta\alpha+\beta_2(\Delta\alpha-s_{\Delta\alpha})k_{\Delta\alpha}+\beta_3\Delta L+\beta_4(\Delta L-s_{\Delta L})k_{\Delta L}+\beta_0$$

where $\Delta\alpha$ is the angle between the previous and next saccade, ΔL is the amplitude difference, β_{1-4} are the slopes of the individual linear segments, $s_{\Delta\alpha}$ is the critical angle, $s_{\Delta L}=0$ and

$k_x = \begin{cases} 1, & x < s_x \\ 0, & \text{else} \end{cases}$. The parameters β_{0-4} and $s_{\Delta\alpha}$ were fitted with a least squares procedure implemented in SciPy 0.9 for each subject individually. Please note that, for visualization purposes, the model fit in Figure 4 was computed by using all of the available data. All other inferences are based on models that were fit on a per-subject basis. We chose a piecewise-linear regression for two reasons: First, the relationship of angle differences and fixation durations seem to exhibit two linear parts (see Figure 4B,D). That is, using a linear regression introduces systematically larger residuals for large angle differences and for small amplitude differences. This is potentially critical because according to our data, changes in slope are not specific to return locations, and thus do not represent a true IOR effect but instead might interact with inferences about effects of IOR. Second, the breakpoint regression is conceptually simple and provides a decent fit with the data ($r^2 = 0.1$, normalized RMSE = 0.16). Analyses that are 'corrected for the saccadic momentum effect' are carried out on the residuals of this regression.

Consecutively, we computed the duration of fixations with respect to the amplitude difference of the previous and consecutive saccade. Figure 4C shows an average over subjects for 30° bins for different saccade amplitude differences. Confidence intervals are based on bootstrapped across subject averages. Figure 4D shows the same but pooled over all subjects and for 1° bins for both angle and amplitude differences. Figure 4E shows the residuals of our piecewise-linear model, that is fixation durations corrected for saccadic momentum. Qualitative inspection shows that little structure remains in the residuals. Specifically, those areas where few samples are available (compare with Figure 2B, top left panel) show larger deviations than those where many samples are available. In an additional analysis (see Text S1) we found that such deviations can be expected even if no effect of angle and amplitude differences is present in the residuals.

Figure 5 shows trials with return saccades aligned to the return fixation and trials without return saccades. Trials without return locations were aligned as follows: For every subject we estimated $P(F_i|L)$ which expresses the probability that a fixation at location i within a trajectory is a return fixation given that the amplitude of the trajectory is L. For every non-return trial we drew a return fixation location from this distribution and aligned the trial to this position. Error bars show bootstrapped 95% confidence intervals.

Feature Analysis

To assess the relationship of return locations and bottom-up saliency, we used a saliency model similar to [30]. We computed 63 different features that are predictors of fixation locations on plain RGB values of the images. We used luminance, saturation, blue/yellow color, and red/green color channels of the stimulus [3]. All features were computed on three different spatial scales, which were created by rescaling the input image with a Gaussian pyramid. For each feature on each spatial scale, we applied three different filters: Gaussian smoothing ($\sigma = 1.1°$), local contrast ($\sigma = 1.1°$), and texture contrast by calculating the local contrast twice on a feature ($\sigma_1 = 1.1°, \sigma_2 = 5.5°$). The local contrast is computed by $C = \sqrt{(I^2 \oplus G) - (I \oplus G)^2}$, where $\oplus$ is the convolution operator and G is a Gaussian kernel with $\mu = 0$ and $\sigma \in \{1.1°, 5.5°\}$.

Additionally, we computed intrinsic dimensionality [31], ID0, ID1, ID2, each with three different kernel sizes (0.12°, 0.52°, 1°), phase-congruency, and phase-symmetry [32,33] as features. We furthermore considered several interactions of these features. We subtracted red/green contrast, blue/yellow contrast, saturation, and saturation contrast (all finest spatial scale) from phase-congruency and symmetry. Concerning intrinsic dimensionality, we compute $ID0_{0.25°} - ID0_{1°}$, $ID0_{1°} - ID2_{0.52°}$, $ID2_{1°}$ – red/green, $ID2_{1°}$-saturation, $ID2_{0.12°}$ – phase congruency. Together with the two last interactions, red-green contrast - saturation contrast and luminance contrast - saturation contrast, this yields 63 different features. Each feature map for each image was z-scored before it was used for further analysis.

To quantify how well a feature can predict fixations and return locations, we used the area under the receiver-operating characteristics curve (AUC). In short, the AUC assesses how well fixations can be separated from control locations on the same image based on the value of a feature at those locations [17,18]. For every feature, we computed the AUC for separating normal fixation locations from control locations and the AUC for separating return locations from control locations. Control locations were chosen from the distribution of fixations on other images, which ensured that control locations follow the spatial distribution of fixations but were not actually fixated locations. We estimated the variability in the data by repeatedly (N = 150) computing both AUCs based on 1000 randomly sampled fixation and control locations. Confidence intervals were subsequently bootstrapped (N = 2000) on these 150 AUC values for each feature. The dependence between patterns of AUC values was well described by a linear relationship (natural scenes: $r^2 = 0.76$, urban scenes: $r^2 = 0.95$). Figure 6D shows the AUC value of every feature for urban and natural scenes with bootstrapped CIs.

To further investigate the relationship between saliency and return locations, we assigned a saliency score to fixations and return locations. A saliency score was obtained by optimally combining features linearly to separate fixations (or return locations) from control locations. The weights for this combination were estimated with a logistic regression that tried to separate fixations from control locations based upon the 63 features. We used feature values at fixated locations as positive samples and feature values at control locations that were fixated on other-images as negative samples for the logistic regression. To test the hypothesis that return locations are more salient than normal fixations, we estimated two-saliency models and assessed how well return-locations can be predicted in comparison to normal fixations. The two models differ with respect to the samples used for training. The return-saccade (RS) model uses only return locations as positive samples, while the fixation (FIX) model uses only fixation locations from trials where no return saccade occurred. Both models were trained repeatedly by splitting the available data into test and training sets. We used leave-one-out cross-validation, where each subject was used for testing once and was not used for training in this run, this ensured that training and test data was completely independent. Both models predicted return locations and normal fixations separately. We found that return locations could be predicted with an average AUC of 0.73 (RS: 0.724, FIX: 0.731) compared to an AUC of 0.67 (RS: 0.667, FIX: 0.674) for normal fixations. A two-way analysis of variance with factors 'model type' and 'fixation type' revealed that both main effects and the interaction between the two are significant ($p < 0.0001$).

Fixation Sampling Strategies

To compute an internal priority map for a given subject and image, we computed a 2D histogram of fixation locations of all other subjects that did not make a return saccade on the same image from the same dataset. To obtain a density map, we convolved this histogram with a Gaussian kernel with full-width-

half-maximum = 1° and normalized the filtered histogram to unit area.

To evaluate the likelihood that a trajectory is drawn from an internal priority map, we interpreted the internal priority map as cell probabilities for a multinomial distribution. How often a location is fixated gives the counts for each cell. The probability of a trajectory is then given by

$$P(x|m) = multinom(x_1,...,x_n; m_1,...,m_n)$$

where x_i encodes the number of fixations for location i, and m_i is the probability of the internal priority map at location i. Subsequently, we compared two different trajectories. In one, the return location is fixated twice, but the last fixation is omitted. In another, the return location is fixated only once, but the last fixation is not omitted. These trajectories differ only in how often the return location and the last fixation are fixated. Thus, the entire comparison amounts to a comparison of internal priority map values at the return location and the last fixation of the trajectory. However, $P(x_{rs}|m) \geq P(x_{\neg rs}|m)$ is only fulfilled when the priority map value for the return saccade is at least twice as large as the value for the last fixation.

Supporting Information

Figure S1 Confidence intervals for the hypothesis that no angle and amplitude effect is present in the residuals of the piecewise-linear model. A shows the upper 97.5% confidence boundary as a function of amplitude and angle differences. Values are larger where fewer samples are available. B shows the percentile of the residuals of the piecewise-linear model in the bootstrap distribution. C shows the lower 2.5% confidence boundary. Values are smaller where fewer samples are available.
(EPS)

Text S1 Text S1 describes how the distribution of samples available for different amplitude and angle difference combinations potentially influences fixation duration estimates.
(PDF)

Author Contributions

Conceived and designed the study: NW PK. Analyzed the data: NW SH NS PK. Contributed reagents/materials/analysis tools: NW SH NS. Wrote the paper: NW SH NS PK.

References

1. Posner MI, Cohen Y (1984) Components of Visual Orienting. Attention and performance X 32: 531–556.
2. Klein RM, MacInnes WJ (1999) Inhibition of return is a foraging fascilitator in visual search. Psychological science 10: 346–352.
3. Itti L, Koch C (2001) Computational modelling of visual attention. Nature reviews Neuroscience 2: 194–203.
4. Parkhurst D, Law K, Niebur E (2002) Modeling the role of salience in the allocation of overt visual attention. Vision research 42: 107–123.
5. Peters RJ, Iyer A, Itti L, Koch C (2005) Components of bottom-up gaze allocation in natural images. Vision Research 45: 2397–2416.
6. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) SUN: A Bayesian framework for saliency using natural statistics. Journal of Vision 8: 32.
7. Zelinsky GJ (2008) A theory of eye movements during target acquisition. Psychological review 115: 787–835.
8. Hooge ITC, Over E a B, van Wezel RJ a, Frens M a (2005) Inhibition of return is not a foraging facilitator in saccadic search and free viewing. Vision research 45: 1901–1908.
9. Smith TJ, Henderson JM (2009) Facilitation of return during scene viewing. Visual Cognition 17: 1083–1108.
10. Smith TJ, Henderson JM (2011) Looking back at Waldo: Oculomotor inhibition of return does not prevent return fixations. Journal of Vision 11: 1–11.
11. Smith TJ, Henderson JM (2011) Does oculomotor inhibition of return influence fixation probability during scene search? Attention, perception & psychophysics 73: 2384–2398.
12. Hooge IT, Frens M a (2000) Inhibition of saccade return (ISR): spatio-temporal properties of saccade programming. Vision research 40: 3415–3426.
13. Anderson AJ, Yadav H, Carpenter RHS (2008) Directional prediction by the saccadic system. Current biology 18: 614–618.
14. Açık A, Sarwary A, Schultze-Kraft R, Onat S, König P (2010) Developmental Changes in Natural Viewing Behavior: Bottom-Up and Top-Down Differences between Children, Young Adults and Older Adults. Frontiers in psychology 1: 207.
15. Kaspar K, König P (2011) Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation. PloS one 6: e21719.
16. Kaspar K, König P (2011) Viewing behavior and the impact of low-level image properties across repeated presentations of complex scenes. Journal of Vision 11: 1–29.
17. Wilming N, Betz T, Kietzmann TC, König P (2011) Measures and Limits of Models of Fixation Selection. PloS one 6: e24038.
18. Tatler BW, Baddeley RJ, Gilchrist ID (2005) Visual correlates of fixation selection: effects of scale and time. Vision research 45: 643–659.
19. Bays P, Husain M (2012) Active inhibition and memory promote exploration and search of natural scenes. Journal of Vision 12: 1–18.
20. Lynch JC, Tian J-R (2006) Cortico-cortical networks and cortico-subcortical loops for the higher control of eye movements. Progress in brain research 151: 461–501.
21. Bisley JW (2011) The neural basis of visual attention. The Journal of physiology 589: 49–57.
22. Farrell S, Ludwig CJH, Ellis L a, Gilchrist ID (2010) Influence of environmental statistics on inhibition of saccadic return. Proceedings of the National Academy of Sciences of the United States of America 107: 929–934.
23. Bahill A, Clark M, Stark L (1975) The main sequence, a tool for studying human eye movements. Mathematical Biosciences 204: 191–204.
24. Ludwig CJH, Farrell S, Ellis L a, Gilchrist ID (2009) The mechanism underlying inhibition of saccadic return. Cognitive psychology 59: 180–202.
25. Hall NJ, Colby CL (2011) Remapping for visual stability. Philosophical transactions of the Royal Society of London Series B, Biological sciences 366: 528–539.
26. Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience 2: 79–87.
27. Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. Nature 434: 387–391.
28. Tatler BW, Vincent BT (2008) Systematic tendencies in scene viewing. Journal of Eye Movement Research 2: 5.
29. Torralba A, Oliva A, Castelhano MS, Henderson JM (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychological review 113: 7
30. Betz T, Kietzmann T, Wilming N, König P (2010) Investigating task-dependent top-down effects on overt visual attention. Journal of Vision 10: 1–14.
31. Saal H (2010) Intrinsic Dimensionality of Visual Stimuli at the Center of Gaze. In: Kühnberger K-U, König P, Walter S, editors. Publications of the Institute of Cognitive Science. Volume 12. 54 pp.
32. Kovesi P (1999) Image features from phase congruency. Videre: Journal of Computer Vision Research 1: 1–26.
33. Kovesi P (2003) Phase congruency detects corners and edges. In: Proceedings of The Australian Pattern Recognition Society Conference. Sydney, USA: IEEE. pp. 309–318.

Publications

Journal Papers

- K. Nakajima, N. M. Schmidt, R. Pfeifer (2015) "Measuring information transfer in a soft robotic arm." *Bioinspir. Biomim.*, 10(3):035007. doi:10.1088/1748-3190/10/3/035007
- N. Wilming, S. Harst, N. M. Schmidt, P. König (2013) "Saccadic Momentum and Facilitation of Return Saccades Contribute to an Optimal Foraging Strategy." *PLoS Comput Biol* 9(1):e1002871. doi:10.1371/journal.pcbi.1002871
- N. M. Schmidt, M. Hoffmann, K. Nakajima, R. Pfeifer (2013) "Bootstrapping perception using information theory: case study in a quadruped robot running on different grounds." *Advs. Complex Syst.* 16:02n03. doi:10.1142/S0219525912500786
- N. M. Schmidt, B. Blankertz, M. S. Treder (2012) "Online detection of error-related potentials boosts the performance of mental typewriters." *BMC Neuroscience*, 13:19. doi:10.1186/1471-2202-13-19
- M. S. Treder, N. M. Schmidt, B. Blankertz (2011) "Gaze-independent high-accuracy brain-computer interfaces based on covert attention and feature attention." *J Neural Eng*, 8:6. doi:10.1088/1741-2560/8/6/066003
- M. S. Treder, A. Bahramisharif, N. M. Schmidt, M. A. J. van Gerven, B. Blankertz (2011) "Posterior alpha modulation induced by covert attention shifts as an input modality for gaze-independent EEG-based brain-computer interfaces." *J Neuroeng Rehabil*, 8:24. doi:10.1186/1743-0003-8-24

Conference Papers

- N. M. Schmidt, M. Baumgarnter, R. Pfeifer (2014) "Actor-Critic Design using Echo State Networks in a Simulated Quadruped Robot." *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. September 14–18, 2014, Chicago, IL, USA. doi:10.1109/IROS.2014.6942862
- M. Hoffmann, N. M. Schmidt, R. Pfeifer, A. K. Engel, A. Maye (2012) "Using Sensorimotor Contingencies for Terrain Discrimination and Adaptive Walking Behavior in the Quadruped Robot Puppy." In T. Ziemke, C. Balkenius, and J. Hallam, Eds. *From Animals to Animats 12, 12th International Conference on Simulation of Adaptive Behavior (SAB)*, August 27–30, 2012, Odense, Denmark, p. 54–64. doi:10.1007/978-3-642-33093-3_6
- G. Kootstra, N. Wilming, N. M. Schmidt, M. Djurfeldt, D. Kragic, P. König (2012) "Learning and Adaptation of Sensorimotor Contingencies: Prism-Adaptation, a Case Study." In T. Ziemke, C. Balkenius, and J. Hallam, Eds. *From Animals to Animats 12, 12th International Conference on Simulation of Adaptive Behavior (SAB)*, August 27–30, 2012, Odense, Denmark, p. 341–350. doi:10.1007/978-3-642-33093-3_34
- M. Hoffmann, N. M. Schmidt, K. Nakajima, F. Iida, R. Pfeifer (2011). "Perception, motor learning, and speed adaptation exploiting body dynamics: Case studies in a quadruped robot." *Proceedings of the 5th International Symposium on Adaptive Motion of Animals and Machines (AMAM)*, 11-14 October 2011, Hyogo, Japan, p. 39-40
- N. M. Schmidt, B. Blankertz, M. S. Treder (2011) "Online Detection of Error-Related Potentials Boosts the Communication Speed of Visual Spellers." *Proceedings of the 2011 5th International Brain-Computer Interface Conference*, September 22–24, 2011, Graz, Austria, p. 208–211
- B. Blankertz, N. M. Schmidt, M. S. Treder (2010) "Gaze-independent BCI spellers based on covert attention and feature attention." *3rd International Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL)*, November 7–10, 2010, Rome, Italy. doi:10.1109/ISABEL.2010.5702885
- N. M. Schmidt, B. Blankertz, M. S. Treder (2010) "Alpha-modulation induced by covert attention shifts as a new input modality for EEG-based BCIs." *Proceedings of the 2010 IEEE Conference on Systems Man and Cybernetics*. October 10–13, 2010, Istanbul, Turkey, p. 481–487. doi:10.1109/ICSMC.2010.5641967